

**OPTIMASI NAÏVE BAYES BERBASIS PSO UNTUK
ANALISIS SENTIMEN TERHADAP PEMBANGUNAN
IBU KOTA NUSANTARA PADA TWITTER**

Oleh

ARIF NUR ALAM

T3119112

SKRIPSI

**Untuk Memenuhi Salah Satu Syarat Ujian
Guna Memperoleh Gelar Sarjana**



**PROGRAM SARJANA
TEKNIK INFORMATIKA
UNIVERSITAS ICHSAN GORONTALO
GORONTALO
2024**

PERSETUJUAN SKRIPSI

OPTIMASI NAÏVE BAYES BERBASIS PSO UNTUK ANALISIS SENTIMEN TERHADAP PEMBANGUNAN IBU KOTA NUSANTARA PADA TWITTER

Oleh

ARIF NUR ALAM

T3119112

SKRIPSI

Untuk memenuhi salah satu syarat ujian

guna memperoleh gelar Sarjan

Program Studi Teknik Informatika.

Ini telah disetujui oleh Tim Pembimbing

Gorontalo, Juni 2024

Pembimbing I

Irma Surya Kumala Idris, M.Kom
NIDN: 0921128801

Pembimbing II

Sudirman S. Panna, M.Kom
NIDN: 0924038205

PENGESAHAN SKRIPSI

OPTIMASI NAÏVE BAYES BERBASIS PSO UNTUK ANALISIS SENTIMEN TERHADAP PEMBANGUNAN IBU KOTA NUSANTARA PADA TWITTER

Oleh

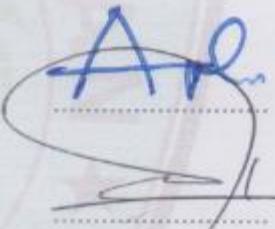
ARIF NUR ALAM

T3119112

Diperiksa oleh Panitia Ujian Strata Satu (S1)

Universitas Ichsan Gorontalo

1. Ketua Pengudi
Amiruddin, M.Kom
2. Anggota
Sunarto Taliki, M.Kom
3. Anggota
Ir. Misrawaty A. Puspa, M.Kom
4. Pembimbing I
Irma Surya Kumala Idris, M.Kom
5. Pembimbing II
Sudirman S. Panna, M.Kom

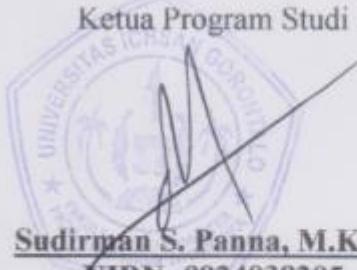



Dekan Fakultas Ilmu Komputer



Irvan A. Salih, M.Kom
NIDN: 0928028101

Ketua Program Studi



Sudirman S. Panna, M.Kom
NIDN: 0924038205

PERNYATAAN SKRIPSI

Dengan ini saya menyatakan bahwa :

1. Karya tulis (Skripsi) saya ini adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (sarjana) baik di Universitas Ichsan Gorontalo maupun di perguruan tinggi lainnya.
2. Karya tulis (Skripsi) saya ini adalah murni gagasan, rumusan, dan penelitian saya sendiri, tanpa bantuan pihak lain, kecuali arahan Tim Pembimbing.
3. Dalam karya tulis (Skripsi) saya ini tidak terdapat karya atau pendapat yang telah dipublikasikan orang lain, kecuali secara tertulis dicantumkan sebagai acuan/situs dalam naskah dan dicantumkan pula dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya tulis ini, serta sanksi lainnya sesuai dengan norma-norma yang berlaku di Universitas Ichsan Gorontalo.

Gorontalo, Juni 2024

Yang Membuat Pernyataan

Materai 10.000

Arif Nur Alam

ABSTRACT

ARIF NUR ALAM. T3119112. THE PSO-BASED NAÏVE BAYES OPTIMIZATION FOR SENTIMENT ANALYSIS ON THE NUSANTARA CAPITAL DEVELOPMENT ON TWITTER

The relocation of Indonesia's Capital City to East Kalimantan, Nusantara Capital City, has triggered various public reactions expressed through social media. This research aims to optimize the Naïve Bayes algorithm using Particle Swarm Optimization (PSO) in sentiment analysis towards the Nusantara Capital City Development on the Twitter platform. Naïve Bayes is used as a text classification algorithm to categorize tweets into positive and negative sentiments, while PSO is used to improve the classification accuracy. Tweet data collection related to the Nusantara Capital City development reveals diverse public opinions. The test results obtained by the Naïve Bayes algorithm indicate an accuracy value of 66% and the addition of Particle Swarm Optimization (PSO) feature selection resulted in an increase in the accuracy value to 70%, indicating an increase in accuracy of 4% from the previous results with 80% data division for training and 20% for testing. This research provides in-depth insight into public perceptions of this major national project, which can be a reference for policymakers and future research.

Keywords: *Naïve Bayes, Particle Swarm Optimization, sentiment analysis, Twitter, Nusantara Capital City*

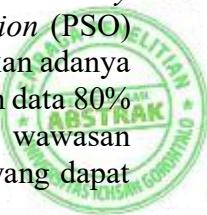


ABSTRAK

ARIF NUR ALAM. T3119112. OPTIMASI NAÏVE BAYES BERBASIS PSO UNTUK ANALISIS SENTIMEN TERHADAP PEMBANGUNAN IBU KOTA NUSANTARA PADA TWITTER

Pemindahan Ibu Kota Indonesia ke Kalimantan Timur, yang dinamai Nusantara, telah memicu berbagai reaksi publik yang diekspresikan melalui media sosial. Penelitian ini bertujuan untuk mengoptimalkan algoritma Naïve Bayes menggunakan *Particle Swarm Optimization* (PSO) dalam analisis sentimen terhadap pembangunan Ibu Kota Nusantara di platform Twitter. Naïve Bayes digunakan sebagai algoritma klasifikasi teks untuk mengkategorikan tweet menjadi sentimen positif dan negatif, sedangkan PSO digunakan untuk meningkatkan akurasi klasifikasi tersebut. Melalui pengumpulan data tweet terkait pembangunan Ibu Kota Nusantara, analisis ini mengungkapkan opini publik yang beragam. Hasil pengujian yang didapatkan oleh algoritma Naïve Bayes memiliki nilai *Accuracy* sebesar 66%, dan penambahan seleksi fitur *Particle Swarm Optimization* (PSO) menghasilkan peningkatan nilai *Accuracy* menjadi 70% yang menunjukkan adanya peningkatan akurasi sebesar 4% dari hasil sebelumnya dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian. Penelitian ini memberikan wawasan mendalam tentang persepsi publik terhadap proyek besar nasional ini, yang dapat menjadi referensi bagi pengambil kebijakan dan penelitian selanjutnya.

Kata Kunci : Naïve Bayes, *Particle Swarm Optimization*, Analisis Sentimen, Twitter, Ibu Kota Nusantara



KATA PENGANTAR

Alhamdulillah, penulis dapat menyelesaikan skripsi ini dengan judul: **“OPTIMASI NAÏVE BAYES BERBASIS PSO UNTUK ANALISIS SENTIMENT TERHADAP PEMBANGUNAN IBU KOTA NUSANTARA PADA TWITTER”**. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan Program Studi S1 di Jurusan Teknik Informatika, Fakultas Ilmu Komputer, Universitas Ichsan Gorontalo.

Penulis sepenuhnya menyadari bahwa skripsi ini tidak mungkin terwujud tanpa ada bantuan dan dorongan dari beberapa pihak, baik bantuan moril maupun materil. Maka dari itu, dengan segala keikhlasan dan kerendahan hati, penulis mengucapkan terima kasih sebesar-besarnya dan penghargaan setinggi-tingginya kepada:

1. Ibu Dr. Hj Juriko Abdussamad, M,Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan Abraham Salihi S.Kom.,M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Gorontalo;
4. Bapak Sudirman Melangi S.Kom.,M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer Universitas Gorontalo;
5. Ibu Irma Surya Kumala Idris, S.Kom.,M.Kom, selaku Wakil Dekan II Bidang Administrasi Umumdan Keuangan Fakultas Ilmu Komputer Universitas Gorontalo;
6. Bapak Sudirman S. Panna, S.Kom.,M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Gorontalo;
7. Ibu Irma Surya Kumala Idris, S.Kom.,M.Kom, selaku Pembimbing I; Utama Yang telah banyak membantu peniliti dalam meyelesaikan penulisan penyusunan penelitian ini.

8. Bapak Sudirman S. Panna, S.Kom.,M.Kom, selaku Pembimbing II; Pendamping yang telah banyak membantu penyusunan penelitian ini dari awal hingga akhir.
9. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang telah berusaha mendidik dan mengajarkan berbagai disiplin ilmu kepada kami,
10. Terkhusus Kedua Orang Tua saya tercinta, atas segala kasih sayang, jerih payah, do'a restu serta telah mendukung penulis mencapai cita-cita untuk meraih gelar sarjana
11. Keluarga dan Saudara saya yang tak henti-hentinya memberikan motivasi dan membantu kepada penulis;
12. Rekan-rekan seperjuangan *Informatic Enginnering* 2019 dan senior-senior saya yang telah banyak membantu dan memberikan dukungan moril kepada penulis;
13. Kepada semua pihak yang ikut membantu dalam penyelesaian skripsi ini yang tak sempat penulis sebutkan satu-persatu.

Semoga Allah SWT melimpahkan balasan atas jasa-jasa mereka kepada kami. Penulis sepenuhnya menyadari bahwa apa yang telah dicapai saat ini masih jauh dari kesempurnaan dan kekurangan. Oleh karena itu penulis sangat mengharapkan adanya kritik dan saran yang konstruktif. Penulis berharap semoga hasil yang telah dicapai ini dapat bermanfaat bagi semua orang, Aamiin.

Gorontalo, Juni 2024

Arif Nur Alam

DAFTAR ISI

PERSETUJUAN SKRIPSI	i
PENGESAHAN SKRIPSI	ii
PERNYATAAN SKRIPSI	iii
ABSTRAK	v
ABSTRACT	v
KATA PENGANTAR	vi
DAFTAR ISI	viii
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiii
BAB I	1
PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	4
1.3. Batasan Masalah	4
1.4. Rumusan Masalah.....	4
1.5. Tujuan Penelitian.....	4
1.6. Manfaat Penelitian.....	4
____1.6.1. Manfaat Teoritis	4
____1.6.2. Manfaat Praktis.....	5
BAB II	6
LANDASAN TEORI.....	6
2.1 Tinjauan Studi.....	6
2.2 Tinjauan Pustaka.....	8
2.2.1 Analisis Sentimen	8
2.2.2 Twitter dan Twitter API.....	9
2.2.3 Klasifikasi	11
2.2.4 Pre-Processing	12
2.2.5 Text Mining.....	13
2.2.6 Python	13

2.2.7	<i>Particle Swarm Optimization (PSO)</i>	13
2.2.8	Naïve Bayes.....	14
2.2.9	Confusion Matrix.....	14
2.3	Kerangka Pikir	17
BAB III	18
METODE PENELITIAN		18
3.1	Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian	18
3.2	Pemodelan Penelitian	19
3.2.1	Pengumpulan Data	20
3.2.2	Proses Klasifikasi	22
3.2.3	Proses Optimasi.....	22
3.2.4	Evaluasi.....	23
3.2.5	Hasil	23
BAB IV	24
PEMBANGUNAN MODEL		24
4.1	Pengumpulan Data.....	24
4.2	Pre-processing.....	25
1)	Cleansing	25
2)	Case folding.....	27
3)	Tokenizing	29
4)	Stopword Removal.....	31
5)	Steaming	33
6)	Normalization.....	35
4.3	Pelabelan Data	37
1.)	Translation.....	37
2.)	Pelabelan Data.....	40
4.4	Hasil Pembobotan TF IDF	42
4.5	Klasifikasi	44
4.6	Pembagian Data training dan Testing.....	45
BAB V	50	
HASIL DAN PEMBAHASAN		50
5.1	Evaluasi Model	50

5.1.1	Naïve Bayes Tanpa Particle Swarm Optimization (PSO)	50
5.1.2	Evaluasi Model Naïve Bayes Dengan Particle Swarm Optimization (PSO)	51
5.2	Analisis Hasil Akurasi.....	60
5.3	Analisa dengan Penelitian Terdahulu Error! Bookmark not defined.	
BAB VI	62
PENUTUP	62
6.1	Kesimpulan	62
6.2	Saran	62
DAFTAR PUSTAKA	64

DAFTAR GAMBAR

Gambar 2.1 Alur Pre-Processing.....	12
Gambar 2.2 Kerangka Pikir	17
Gambar 3.1 Pemodelan yang diusulkan	19
Gambar 4.1 <i>Source Code</i> Tahapan <i>Cleaning</i>	25
Gambar 4.2 <i>Source code</i> tahapan <i>Case Folding</i>	27
Gambar 4.3 <i>Source code</i> tahapan <i>Tokenizing</i>	29
Gambar 4.4 <i>Source code</i> Stopword Removal	31
Gambar 4.5 <i>Source code Steamming</i>	33
Gambar 4.6 <i>Source code Normalization</i>	35
Gambar 4.7 <i>Source code Translation</i>	38
Gambar 4.8 <i>Source code Pelabelan</i>	40
Gambar 4.9 Viusalisasi Sentimen.....	42
Gambar 4.10 <i>Source code</i> TF-IDF	43
Gambar 4.11 Visualisasi top 20 term	43
Gambar 4.12 Visualisasi Word Cloud TF-IDF.....	44
Gambar 4.13 <i>Source code</i> Pembagian data latih dan data uji	45
Gambar 4.14 <i>Source code</i> Naïve Bayes tanpa PSO	46
Gambar 4.15 <i>Source code</i> Naïve Bayes dengan PSO.....	49
Gambar 5.1 Visualisasi Confusion Matrix: Naive Bayes tanpa PSO	50
Gambar 5.2 Visualisasi Population Size 1-10	53
Gambar 5.3 Visualisasi Maximum Number Of Generation	55
Gambar 5.4 Visualisasi Inertia Weight (omega)	57
Gambar 5.5 Visualisasi Confusion Matrix: Naive Bayes dengan PSO	58
Gambar 5.6 Grafik dari Perbandingan Nilai Accuracy.....	60

DAFTAR TABEL

Tabel 2.1	: Penelitian Terdahulu	6
Tabel 2.2	: Confusion Matrix.....	15
Tabel 3.1	: Sampel Cleansing	20
Tabel 3.2	: Sampel Case Folding	20
Tabel 3.3	: Sampel Tokenizing.....	21
Tabel 3.4	: Sampel Stopword Removal	21
Tabel 3.4	: Sampel Stemming	22
Tabel 3.6	: Sampel Normalization.....	22
Tabel 4.1	: Data <i>Crawl</i>	24
Tabel 4.2	: Data Cleaning	26
Tabel 4.3	: Data Case Folding.....	28
Tabel 4.4	: Data Tokenizing.....	30
Tabel 4.5	: Data Stopword Removal	32
Tabel 4.6	: Data Stemming	34
Tabel 4.7	: Data Normalization.....	36
Tabel 4.8	: Data Translation.....	38
Tabel 4.9	: Data Pelabelan	40
Tabel 4.10	: Data <i>Training</i>	46
Tabel 4.11	: Data <i>Testing</i>	47
Tabel 4.12	: Hasil hitung probabilitas kata Doc 11	47
Tabel 5.1	: Hasil Eksperimen Algoritma Naïve Bayes berbasis PSO dengan nilai Parameter Population Size 1-10	52
Tabel 5.2	: Hasil Eksperimen Algoritma Naïve Bayes berbasis PSO dengan nilai Parameter Maximum Number of Generation 10- 100 atau berkelipatan 10	54
Tabel 5.3	: Hasil Eksperimen Algoritma Naïve Bayes berbasis PSO dengan nilai parameter Inertia Weight 1.0 - 2.0	56
Tabel 5.4	: Hasil pada Algoritma Naïve Bayes berbasis PSO	58

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pemindahan Ibu Kota Negara Indonesia mulai dibahas oleh Pemerintah Indonesia pada tahun 2019. Melalui rapat terbatas yang diselenggarakan oleh Pemerintah Indonesia, Presiden Indonesia menetapkan pemindahan Ibu Kota Negara (IKN) ke luar Pulau Jawa yang tertulis dalam Rencana Pembangunan Jangka Menengah Nasional Tahun Anggaran 2020-2024 [1]. Pada rapat panitia khusus Rancangan Undang-Undang Ibu Kota Negara (RUU IKN) yang dilaksanaan pada tanggal 17 Januari 2022, Kepala Bappenas mengumumkan bahwa Ibu Kota baru di Kalimantan Timur akan diberi nama "Nusantara" [1]. Mengenai pemindahan Ibu Kota Nusantara, yakni pada masa Presiden Bapak Ir. H. Joko Widodo, Pembangunan Ibu Kota Nusantara pada pertengahan Maret 2022 akan direlokasi secara bertahap dari tahun 2024 hingga 2045. Tepat pada Senin, 26 Agustus 2019, Melalui siaran pers, lokasi baru IKN berada di Kabupaten Penajam Paser Utara Provinsi Kalimantan Timur, tepatnya di sebagian Kabupaten Kutai Kartanegara [2].

Menurut undang-undang no 3 tahun 2022 pasal 1 ayat 13 yang berbunyi: rencana induk Ibu Kota Nusantara adalah dokumen perencanaan terpadu dalam melaksanakan persiapan, pembangunan, dan pemindahan Ibu Kota Negara, serta penyelenggaraan pemerintahan daerah khusus Ibu Kota Nusantara. Adapun visi pemindahan Ibu Kota dijelaskan dalam pasal 2 yang berbunyi: Ibu Kota Nusantara memiliki visi sebagai kota dunia untuk semua yang dibangun dan dikelola dengan tujuan untuk : menjadi kota berkelanjutan di dunia, sebagai penggerak ekonomi Indonesia di masa depan, dan menjadi simbol identitas nasional yang merepresentasikan keberagaman Bangsa Indonesia, berdasarkan Pancasila dan UUD Negara Republik Indonesia tahun 1945. Disisi lain menurut portal berita bbc.com rencana pembangunan Ibu Kota Negara (IKN) baru bernama Nusantara di Kalimantan Timur disebut pakar kebijakan publik akan memberatkan pos

Anggaran Pendapatan dan Belanja Negara (APBN) yang terjepit di tengah pengeluaran besar untuk penanganan kesehatan dan pemulihan ekonomi akibat pandemi covid-19 [3]. Menanggapi hal itu banyak opini berkembang di masyarakat tentang pemindahan Ibu Kota. Mengingat Ibu Kota Negara baru di Indonesia merupakan hal yang sensitif sehingga ramai diperbincangkan di media sosial termasuk Twitter. Ada pihak yang setuju ada pula yang tidak. Berdasarkan banyaknya opini masyarakat di Twitter tentang pemindahan Ibu Kota, kita bisa menggunakan analisis sentimen untuk meninjau opini mana yang lebih dominan di masyarakat.

Masyarakat memanfaatkan media sosial sebagai saluran untuk menyampaikan pendapat mereka dalam merespons suatu permasalahan atau fenomena. Salah satu platform media sosial yang paling populer digunakan adalah Twitter. Pengguna media sosial Twitter di Indonesia, hingga Februari 2022, mencapai 58,3% dari jumlah pengguna internet yang berumur 16 sampai 64 tahun, atau hampir 112 juta pengguna [4]. "Media sosial Twitter sering digunakan sebagai sarana promosi dan penyampaian informasi kepada publik. Banyaknya komentar publik yang mencerminkan opini dapat dijadikan sumber informasi yang berharga. Namun, untuk mengoptimalkan pemanfaatan ini, diperlukan analisis yang tepat, seperti analisis sentimen, untuk mendapatkan informasi yang bermanfaat bagi berbagai pihak.

Analisis sentimen adalah suatu teknik yang digunakan untuk menganalisis pendapat, penilaian, sentimen, sikap, evaluasi, serta emosi yang diungkapkan oleh individu terhadap berbagai entitas, seperti produk, organisasi, layanan, masalah, individu, topik, peristiwa, dan atribut yang terkait [4]. Penentuan sentimen positif dan negatif pada tweet dapat dilakukan secara manual, namun dengan pertumbuhan jumlah pengguna yang semakin besar, jumlah opini yang dihasilkan juga meningkat, yang membutuhkan waktu dan usaha yang signifikan. Oleh karena itu, diperlukan alat yang mampu secara otomatis menganalisis dan mengklasifikasikan banyak tweet ke dalam sentimen positif atau negatif. Metode yang efektif untuk analisis sentimen ini adalah Naïve Bayes.

Penelitian tentang analisis sentimen sebelumnya telah banyak dilakukan diantaranya implementasi algoritma Naive Bayes, *Support Vector Machine* dan K-Nearest Neighbors untuk analisa sentimen aplikasi Halodoc [5]. Penelitian selanjutnya menggunakan algoritma Naive Bayes, pada penelitian ini dihasilkan nilai akurasi sebesar 81,07% [6]. Berdasarkan penelitian tersebut dapat dilihat bahwa algoritma Naive Bayes menghasilkan nilai akurasi yang cukup tinggi. Naïve Bayes merupakan salah satu algoritma yang paling banyak digunakan untuk pengklasifikasian suatu dokumen karena kecepatan dan keakuratannya yang tinggi. Namun pada penelitian yang lain algoritma KNN menghasilkan nilai akurasi tertinggi dibandingkan dengan Naive Bayes [2].

Performa metode Naive Bayes meski memiliki akurasi yang baik namun masih dapat ditingkatkan dengan berbagai macam teknik optimasi seperti peningkatan bobot atribut, seleksi atribut, dan penerapan seleksi fitur. Penerapan seleksi fitur dilakukan karena metode Naive Bayes sangat sensitif terhadap seleksi fitur. Pemilihan fitur dapat meningkatkan skalabilitas, efisiensi, dan akurasi dari metode Naive Bayes [7]. *Particle Swarm Optimization* (PSO) adalah algoritma optimasi stokastik berbasis populasi yang dimotivasi oleh kecerdasan perilaku kolektif dari beberapa hewan seperti kawanan burung atau kumpulan ikan. Terdapat beberapa parameter untuk melakukan optimasi dengan menggunakan PSO, di antaranya jumlah partikel yang menggambarkan kelompok data atau swarm, posisi vektor tiap partikel dalam kelompok data, arah gerakan dari partikel (velocity), learning rate, dan learning rate inertia [7]. Penerapan teknik optimasi *Particle Swarm Optimization* mempengaruhi nilai akurasi akhir sebuah algoritma [8]. PSO banyak digunakan untuk memecahkan masalah optimisasi dan sebagai pemecah masalah pemilihan fitur [9].

Berdasarkan permasalahan yang telah disebutkan dalam latar belakang diatas, penulis bertujuan untuk mengetahui perbedaan nilai akurasi yang dihasilkan oleh Naïve Bayes tanpa optimasi dan dengan optimasi menggunakan PSO dalam menilai orientasi sentimen masyarakat terhadap Pembangunan Ibu Kota Nusantara di *twitter*. Hasil pengukuran diharapkan dapat memberikan penilaian akurat yang dapat digunakan sebagai referensi bagi peneliti selanjutnya.

1.2. Identifikasi Masalah

Adapun identifikasi masalah yang bisa didapat dari latar belakang tersebut adalah "Dibutuhkannya Optimasi metode Naïve Bayes berbasis PSO pada analisis sentimen terhadap opini masyarakat terkait topik yang sedang populer yakni Pembangunan Ibu Kota Nusantara di media sosial twitter".

1.3. Batasan Masalah

Agar Penelitian lebih terarah dan tidak menyimpang dari rumusan masalah yang ada, maka Batasan dari penelitian ini adalah :

1. Tweet yang diambil dan dianalisis hanya tweet berbahasa Indonesia.
2. Tweet hanya dikelompokkan kedalam 2 kelas, yaitu positif dan negatif.

1.4. Rumusan Masalah

Berdasarkan Latar Belakang diatas, maka penulis merumuskan masalah yakni Bagaimana kinerja PSO (*Particle Swarm Optimization*) dalam Optimasi Metode Naïve Bayes Pada *Sentiment Analysis* terkait respon masyarakat terhadap Pembangunan Ibu Kota Nusantara di media sosial *twitter*.

1.5. Tujuan Penelitian

Berdasarkan rumusan masalah diatas, maka tujuan penelitian yakni Mengetahui kinerja PSO (*Particle Swarm Optimization*) dalam Optimasi Metode Naïve Bayes Pada *Sentiment Analysis* terkait respon masyarakat terhadap Pembangunan Ibu Kota Nusantara di media sosial twitter.

1.6. Manfaat Penelitian

1.6.1. Manfaat Teoritis

Secara teoritis, penelitian ini dapat menjadi referensi dan menambah wawasan guna perkembangan ilmu pengetahuan dan teknologi, utamanya dalam bidang ilmu komputer, yakni berupa Optimasi metode Naïve Bayes berbasis PSO dalam menganalisis data.

1.6.2. Manfaat Praktis

Dengan tercapainya tujuan penelitian diatas, maka diharapkan akan bermanfaat bagi semua pihak. Adapun manfaat penelitian ini adalah :

1. Dengan menggunakan algoritma *Naïve Bayes* dapat melakukan klasifikasi sentimen terhadap data tweet.
2. Bagi Universitas Ichsan, dapat digunakan sebagai tambahan referensi yang terkait dengan peningkatan pengguna media sosial seiring kemajuan teknologi.
3. Bagi penulis, untuk memenuhi salah satu persyaratan kelulusan strata satu (S1) Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Ichsan Gorontalo.
4. Bagi Pembaca, dengan adanya penulisan ini dapat bermanfaat sebagai sumber pembelajaran untuk penelitian selanjutnya serta menambah wawasan.

BAB II

LANDASAN TEORI

2.1 Tinjauan Studi

Tabel 2. 1 Penelitian Terdahulu

NO	Peneliti	Judul	Tahun	Metode	Deskripsi Singkat
1	Irma Surya Kumala Idris, Yasin Aril Mustofa & Irwan Abraham Salihi	Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma <i>Support Vector Machine</i> (SVM)	2022	<i>Support Vector Machine</i>	Penelitian ini bertujuan untuk mengklasifikasi komentar dari pengguna aplikasi Shopee menjadi komentar positif dan negatif dengan menganalisis ulasan yang diberikan oleh pengguna. Selain itu, penelitian ini juga bertujuan untuk mengevaluasi kinerja metode klasifikasi yang digunakan. Data dalam penelitian ini diambil dari ulasan pengguna aplikasi Shopee dengan menggunakan metode web scraping, yang menghasilkan sekitar 3000 data ulasan. Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine mampu memberikan kinerja yang baik, dengan tingkat akurasi mencapai 98% dan f1-score sebesar 0,98.
2	Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, Fitri Nurapriani	Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma	2023	<i>Naïve Bayes</i> dan <i>KNN</i>	Hasil penelitian menyajikan hasil komparasi kinerja metode tersebut bahwa metode Naïve Bayes memberikan tingkat akurasi analisis sentimen sebesar 82.27%, nilai Precision sebesar

NO	Peneliti	Judul	Tahun	Metode	Deskripsi Singkat
		Naïve Bayes dan KNN			86.36% dan nilai Recall sebesar 76.93%. Kinerja metode KNN juga menyajikan hasil analisis dengan tingkat akurasi sebesar 88,12%, Precision sebesar 93.%98 dan nilai recall sebesar 81.53%. Berdasarkan hasil analisis tersebut maka proses analisis menggunakan metode KKN mengungguli metode NB dalam mengukur sentimen terhadap pemindahan Ibu Kota Nusantara
3	Nur Fitriyah, Budi Warsito, Di Asih I Maruddani	Optimasi Parameter PSO Berbasis SVM Untuk Analisis Sentimen Review Jasa Maskapai Penerbangan Berbahasa Inggris	2020	<i>Support Vector Machine, PSO</i>	PSO mampu mengoptimasikan model SVM dengan nilai akurasi model SVM sebelum diterapkannya seleksi fitur PSO adalah sebesar 84,25% dan setelah diterapkannya seleksi fitur PSO akurasi meningkat menjadi 87,39%. Terdapat kenaikan akurasi sebesar 3.14%
4	Elly Indrayuni, Acmad Nurhadi	Optimasi <i>Naïve Bayes</i> Berbasis PSO Untuk Analisa Sentimen Perkembangan <i>Artificial Intelligence</i> Di Twitter	2023	<i>Naïve Bayes, PSO</i>	Hasil pengujian klasifikasi sentimen terhadap perkembangan Artificial Intelligence menggunakan Naive Bayes menghasilkan nilai akurasi sebesar 86.42%. Sedangkan untuk hasil pengujian klasifikasi sentimen menggunakan Naive Bayes berbasis Particle Swarm Optimization (PSO) mengalami kenaikan dengan nilai akurasi sebesar 87.55%. Hasil penelitian ini yaitu

NO	Peneliti	Judul	Tahun	Metode	Deskripsi Singkat
					penggunaan PSO sebagai teknik optimasi pada algoritma Naive Bayes terbukti sebagai model algoritma terbaik dalam analisa sentimen perkembangan Artificial Intelligence untuk teks berbahasa Inggris.
5	Adam Prasetya, Ferdiansyah, Yesi Novaria Kunang, Edi Surya Negara, Winoto Chandra	Sentimen Analisis Terhadap <i>Cryptocurrency</i> Berdasarkan Comment Dan Reply Pada Platform Twitter	2021	<i>Support Vector Machine (SVM) & Naïve Bayes</i>	Penelitian ini bertujuan untuk melakukan analisis sentimen data dengan melakukan klasifikasi data dari Twitter. Data tersebut akan melalui tahap text mining untuk membersihkan data yang tidak sempurna. Setelah itu, data tweet akan diklasifikasikan ke dalam tiga kategori, yaitu positif, negatif, dan netral.

2.2 Tinjauan Pustaka

2.2.1 Analisis Sentimen

Analisis sentimen adalah bidang studi yang berfokus pada analisis pendapat, sentimen, evaluasi, penilaian, sikap, dan emosi seseorang terhadap berbagai entitas, seperti produk, organisasi, individu, isu konkret, dan peristiwa. Bidang ini memiliki berbagai istilah yang digunakan, seperti analisis sentimen, penambangan opini, ekstraksi opini, penambangan sentimen, yang semuanya masuk dalam cakupan yang lebih luas dari analisis sentimen atau penambangan opini. Dalam dunia industri, istilah “analisis sentimen” sering digunakan, sedangkan di dunia pendidikan, baik “analisis sentimen” maupun “opinion mining” kerap digunakan. Kata “analisis sentimen” pertama kali muncul pada tahun 2003 oleh Nasukawa dan Yi, sementara “opinion mining” pertama kali digunakan pada tahun 2003 oleh Dave Lawrence dan Pennock [10].

Meskipun linguistik dan Pemrosesan Bahasa Alami (Natural Language Processing atau NLP) memiliki sejarah panjang, hanya sedikit penelitian yang dilakukan tentang pendapat dan sentimen manusia sebelum tahun 2000. Sejak saat itu, bidang ini telah menjadi sangat aktif sebagai bidang penelitian. Ada beberapa alasan mengapa demikian. Pertama, analisis sentimen memiliki penerapan yang luas dan relevan di hampir semua bidang. Di industri, perkembangan komersial dalam analisis sentimen telah memberikan dorongan kuat untuk penelitian dalam domain ini. Kedua, analisis sentimen menyediakan berbagai masalah penelitian yang menantang yang belum pernah dipelajari sebelumnya. Ketiga, dalam zaman informasi saat ini, kita memiliki akses ke volume data besar yang berasal dari media sosial di web, dan tanpa data ini, banyak penelitian tidak akan mungkin. Analisis sentimen telah menjadi pusat perhatian dalam penelitian media sosial, dan oleh karena itu, penelitian dalam analisis sentimen memiliki dampak signifikan tidak hanya dalam NLP, tetapi juga dalam manajemen ilmiah, sains politik, ekonomi, dan sains sosial, karena semuanya dipengaruhi oleh beragam pendapat masyarakat [10].

2.2.2 Twitter dan Twitter API

Twitter adalah situs web yang dimiliki dan dioperasikan oleh Twitter Inc., yang menyediakan platform jaringan sosial berupa mikroblog, memungkinkan pengguna untuk mengirim dan membaca pesan yang disebut Tweets. Seperti yang diungkapkan oleh Mollett dan rekan-rekannya, ribuan akademisi dan peneliti dari berbagai tingkat pengalaman dan disiplin ilmu menggunakan Twitter secara aktif setiap hari [11]. Twitter adalah sebuah platform jejaring sosial gratis yang sangat populer, yang memungkinkan pengguna untuk berbagi informasi dalam bentuk umpan berita real-time dengan posting komentar singkat mengenai pengalaman dan pemikiran mereka. Pesan publik yang dikirim dan diterima melalui Twitter, yang dikenal sebagai 'tweet,' dibatasi hingga 140 karakter dan dapat mencakup tautan ke blog, halaman web, gambar, video, dan berbagai jenis konten online lainnya [11].

Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda # (*hashtag*), sementara untuk menyebutkan atau membalas pesan dari pengguna lain dapat menggunakan tanda @. Fitur yang terdapat dalam *Twitter*, antara lain:

1. Halaman Utama (*Home*)

"Pada halaman utama, pengguna dapat melihat Tweet yang diposting oleh orang-orang yang menjadi teman atau yang diikuti (*following*) oleh pengguna tersebut.

2. Profil (*Profile*)

Pada halaman ini, profil dan data diri pengguna, serta semua Tweet yang pernah dibuat, akan dapat dilihat oleh semua orang.

3. Pengikut (*Followers*)

Pengikut (*followers*) adalah pengguna lain yang ingin mengikuti akun pengguna. Ketika pengguna lain mengikuti akun seseorang, maka Tweet dari akun tersebut akan muncul di halaman utama pengguna yang mengikuti akun tersebut.

4. Mengikuti (*Following*)

"Kebalikan dari pengikut adalah *following*, yaitu akun pengguna yang diikuti oleh pengguna lain sehingga *Tweet* yang diposting oleh akun yang diikuti akan muncul di halaman utama pengguna yang melakukan *following*.

5. *Mentions*

Konten ini biasanya berisi balasan dalam percakapan, sehingga pengguna dapat langsung menandai atau menyebut pengguna lain yang ingin diajak berbicara.

6. *Favorite*

Pengguna dapat menandai *Tweets* sebagai favorit agar *Tweets* tersebut mudah ditemukan dan tidak hilang di antara jumlah *Tweets* yang banyak di halaman sebelumnya.

7. Pesan Langsung (*Direct Message*)

Fungsi pesan langsung lebih bisa disebut SMS karena pengiriman pesan langsung di antara pengguna.

8. *Hashtag*

Hashtag '#' yang ditulis sebelum topik tertentu memungkinkan pengguna lain untuk mencari topik yang serupa yang telah dibagikan oleh pengguna lain.

9. List

Pengguna Twitter dapat mengelompokkan pengikut mereka ke dalam daftar (*list*), sehingga memudahkan mereka untuk melihat secara keseluruhan akun pengguna yang mereka ikuti.

10. Topik Terkini (*Trending Topic*)

Topik yang sedang banyak dibicarakan oleh banyak pengguna dalam waktu yang bersamaan disebut sebagai tren atau *trending topic*.

Twitter telah menjadi media sosial yang sangat populer dan efektif. Mereka menyediakan *Application Programming Interface* (API) yang memudahkan pengembang perangkat lunak untuk mengambil data dari Twitter dan mengolahnya. Pengumpulan data dari Twitter dapat digunakan untuk berbagai kebutuhan, termasuk analisis sentimen. Pengunduhan data dari Twitter memerlukan Keys dan Token Twitter API, yang juga terintegrasi dengan banyak perangkat pemrograman, memudahkan para peneliti untuk melakukan analisis opini atau sentimen [11].

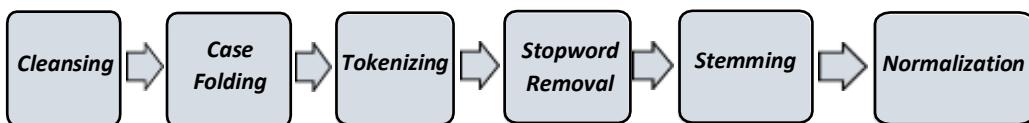
2.2.3 Klasifikasi

Klasifikasi adalah proses untuk menemukan sekumpulan model yang dapat mendeskripsikan dan membedakan kelas data. Tujuan utama klasifikasi adalah memprediksi kelas dari suatu objek yang kelasnya tidak diketahui. Proses klasifikasi melibatkan dua langkah, yaitu pembangunan model klasifikasi menggunakan sekumpulan data pelatihan (*training data set*) yang kelasnya telah didefinisikan sebelumnya, dan penggunaan model ini untuk mengklasifikasikan data uji (*prediction*), kemudian mengukur akurasi dari model yang telah dibangun.

Klasifikasi memiliki berbagai aplikasi, seperti dalam diagnosa medis, pemasaran selektif, pengajuan kredit perbankan, analisis sentimen pada email, dan banyak lagi. Ada berbagai jenis model klasifikasi yang dapat digunakan, termasuk *decision trees*, *naive Bayes classifier*, *k-nearest-neighbor classifier*, jaringan saraf tiruan (*neural network*), dan *Support Vector Machine*".

2.2.4 Pre-Processing

Preprocessing adalah tahap pertama dalam penambangan teks yang bertujuan untuk mengubah data ke dalam format yang diperlukan. Langkah ini diperlukan untuk mencari, mengolah, dan menyusun informasi, serta menganalisis hubungan teks berdasarkan data yang dapat berupa terstruktur maupun tidak terstruktur [12].



Gambar 2.1 Alur Pre-Processing

1. Cleansing

Cleansing merupakan proses pembersihan atau penghapusan kata yang tidak penting. Contoh kata yang tidak penting seperti “@”, “#”, ”rt”, dan lain sebagainya.

2. Case folding

Adalah proses perubahan semua huruf pada dokumen tweet menjadi huruf kecil. Hanya huruf a sampai z yang akan diproses. Karakter selain huruf akan diabaikan.

3. Tokenizing

Tokenizing adalah proses untuk memotong dokumen menjadi pecahan kecil yang dapat berupa bab, sub-bab, paragraph, kalimat, dan kata (token).

4. Stopword Removal

Stopword adalah proses penghilang kata yang tidak berkaitan dengan topik dokumen, karena kata tersebut tidak akan mempengaruhi keakuratan dalam klasifikasi sentimen

5. Stemming

Stemming merupakan proses untuk mencari stem (kata dasar) dari kata hasil stopword removal (filtering). Terdapat dua aturan dalam melakukan

stemming yaitu dengan pendekatan kamus dan pendekatan aturan (Utomo, 2013).

6. Normalization

Normalization merupakan proses untuk mengubah kata yang tidak baku sesuai ejaan.

2.2.5 Text Mining

Text Mining adalah proses ekstraksi pola dari data yang berasal dari teks. Ini adalah sebuah disiplin ilmu yang bergantung pada Information Retrieval, Data Mining, Machine Learning, Statistik, dan Linguistik Komputasi untuk menganalisis dan menggali wawasan dari data teks [13].

2.2.6 Python

Python adalah bahasa pemrograman yang populer digunakan oleh perusahaan besar maupun pengembang untuk membangun aplikasi desktop, web, dan mobile. Peneliti sering memilih Python karena sintaksis kodennya yang jelas dan tingkat fungsionalitas yang tinggi, yang mempermudah mereka dalam menyelesaikan tugas akhir yang melibatkan pemrograman [14].

Python adalah bahasa pemrograman yang relatif mudah dan efisien. Python juga termasuk dalam kategori high-level programming language. Python dirancang dengan perhatian khusus pada kenyamanan pengembang, sehingga mereka dapat bekerja dengan lebih efisien dalam menyelesaikan tugas-tugas mereka.

2.2.7 Particle Swarm Optimization (PSO)

Particle Swarm Optimization (PSO) merupakan teknik metaheuristik berbasis populasi yang terinspirasi dari perilaku kawanan burung dan ikan saat berpindah dari satu tempat ke tempat lain untuk mencari makanan. Menurut (Pandhu Wijaya & Agus Santoso, 2018), Metode Particle Swarm Optimzation (PSO) adalah pendekatan yang berjalan atas dasar pola interaksi yang saling berkaitan dan memperhitungkan faktor kuantitas berupa populasi sebagai dasar utama penelitian dengan data yang besar [15].

Menurut (Li et al., 2019), *Particle Swarm Optimization* (PSO) memiliki algoritma yang sederhana untuk diimplementasikan dan dikonvergen dengan cepat, namun juga dengan mudah jatuh ke optimal lokal. Dalam mengimplementasikan PSO ini melakukan penyesuaian terhadap parameternya. Menurut (Mahapatra et al., 2022), Kinerja dari algortima PSO sangat bergantung dengan strategi parameter yang tepat dalam menyempurnakan paramernya, Inertia Weight merupakan salah satu parameter PSO yang digunakan dalam menyeimbangkan eksplorasi dan eksploitasi PSO [15].

2.2.8 Naïve Bayes

Naïve Bayes Classifier adalah teknik klasifikasi statistik yang memperhitungkan probabilitas kemungkinan kelas, termasuk probabilitas signifikansi kelas. Saat menggunakan metode ini untuk mengkategorikan data twitter, data tersebut akan dikategorikan menurut skor positif dan negatifnya. Pengklasifikasian Naïve Bayes adalah metode klasifikasi berdasarkan teorema bayes. Teorema Bayes adalah teori statistik yang dapat diterapkan untuk mempelajari hipotesis tertentu. Bayes adalah teknik prediksi berbasis probabilistik berdasarkan teorema bayes dengan asumsi independensi yang kuat[14]. Naïve Bayes memiliki ciri utama seperti asumsi dari masing-masing kondisi atau kejadian yang sangat kuat. Metode tersebut digunakan untuk pengklasifikasian sentiment karena mencakup sejumlah keunggulan yang berbeda satu sama lain, berakurasi tinggi, cepat, dan sederhana [14].

2.2.9 Confusion Matrix

Confusion matrix adalah tabel yang digunakan untuk mencatat hasil kinerja dari suatu proses klasifikasi. Dengan menggunakan confusion matrix, hasil klasifikasi dapat diringkas menjadi nilai akurasi dan tingkat kesalahan. Dengan mengetahui berapa jumlah data yang diklasifikasikan dengan benar, kita dapat menghitung akurasi dari prediksi tersebut. Selain itu, dengan mengetahui berapa jumlah data yang diklasifikasikan dengan salah, kita dapat menghitung tingkat kesalahan dari prediksi tersebut. Kedua kuantitas ini digunakan sebagai metrik

kinerja dalam proses klasifikasi.

Tabel 2.2 Confusion Matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive(TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative(TN)

Matriks tersebut memiliki empat nilai yang dijadikan acuan dalam perhitungan, yaitu :

- a) *True Positive* (TP), ketika kelas yang diprediksi positif dan faktanya positif.
- b) *True Negative* (TN), ketika kelas yang diprediksi negatif, dan faktanya negatif.
- c) *False Positive* (FP), ketika kelas yang diprediksi positif dan faktanya negatif.
- d) *False Negative* (FN), ketika kelas yang diprediksi negatif dan faktanya positif.

Berikut beberapa ukuran yang biasa dipergunakan untuk mengevaluasi model klasifikasi.

1. Akurasi : menyatakan persentase dari data uji yang diklasifikasikan dengan benar. Untuk menghitung akurasi digunakan rumus :

$$Akurasi = \frac{\text{jumlah data diprediksi benar}}{\text{jumlah data yang diprediksi}}$$

atau

$$Akurasi = \frac{TP + TN}{(TP + TN + FP + FN)}$$

2. Error rate : menyatakan persentase data uji yang diklasifikasikan dengan tidak benar. Untuk menghitung error rate digunakan rumus :

$$Error\ Rate = \frac{FP + FN}{(FP + FN + TP + TN)}$$

3. Recall : menyatakan persentase data positif yang diklasifikasikan dengan benar. Untuk menghitung recall digunakan rumus :

$$Recall = \frac{TP}{TP + FN}$$

4. Specificity : menyatakan persentase data negatif yang diklasifikasikan dengan benar. Untuk menghitung specificity digunakan rumus :

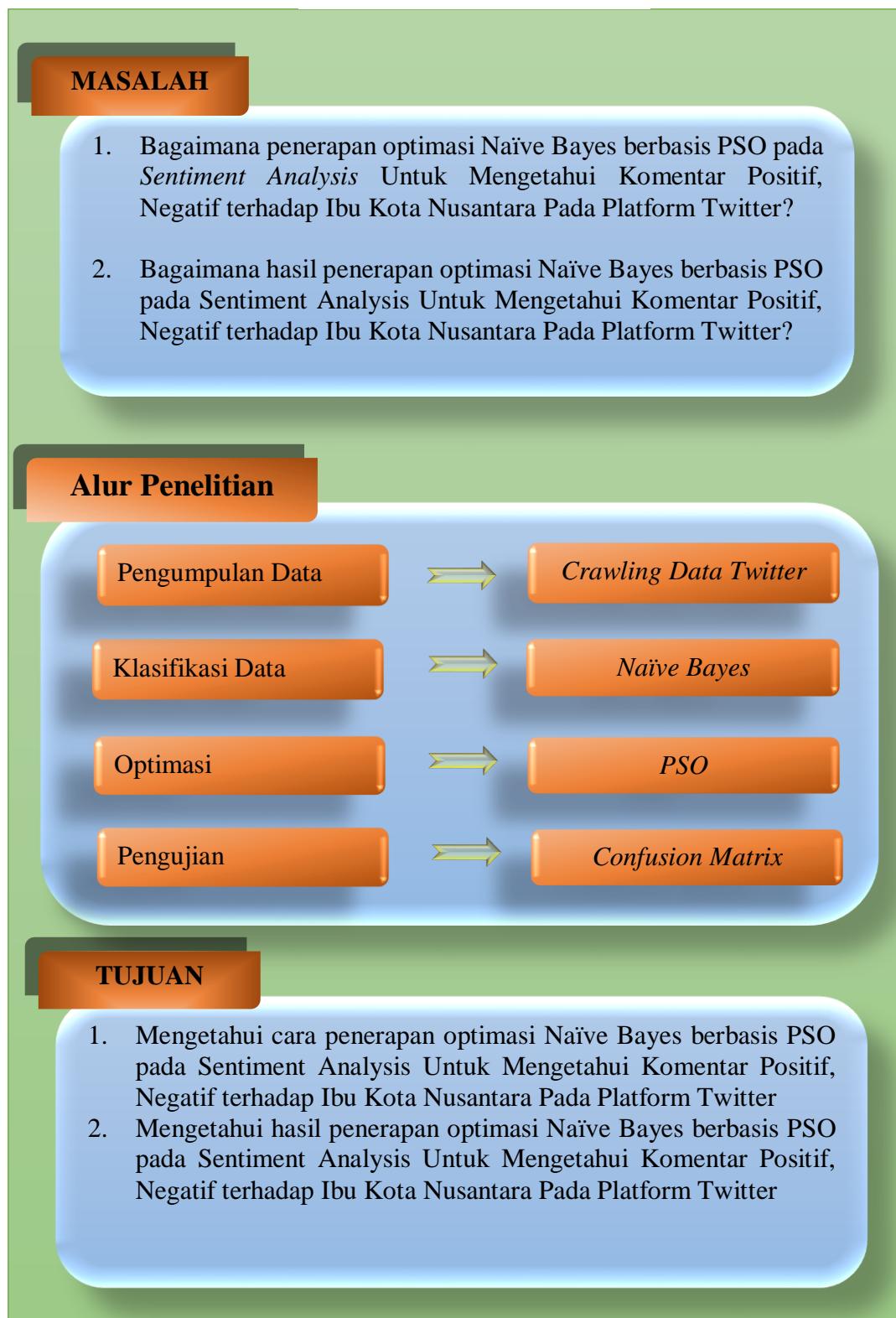
$$Specificity = \frac{TN}{FP + TN}$$

5. Precision : beberapa persentase data yang dilabeli positif adalah memang benar positif. Untuk menghitung precision digunakan rumus :

$$Precision = \frac{TP}{TP + FP}$$

2.3 Kerangka Pikir

Gambar 2.2 kerangka pikir



BAB III

METODE PENELITIAN

3.1 Jenis, Metode, Subjek, Objek, Waktu, dan Lokasi Penelitian

3.1.1 Jenis dan Metode

Penelitian ini termasuk dalam jenis penelitian kuantitatif, yang melibatkan analisis data berbasis text dan perhitungan statistik untuk menguji hipotesis. Penelitian ini memerlukan penggunaan alat analisis dan desain penelitian yang khusus. Berdasarkan tujuannya, penelitian ini dapat dikategorikan sebagai penelitian terapan, yang fokus pada pemecahan masalah praktis. "Penelitian ini bertujuan untuk menjawab pertanyaan spesifik yang berkaitan dengan pengembangan kebijakan, pengambilan tindakan, atau perbaikan kinerja tertentu. Hasil dari penelitian ini umumnya berupa informasi yang diperlukan untuk pengambilan keputusan dalam menyelesaikan masalah-masalah praktis.

3.1.1 Subject dan Object

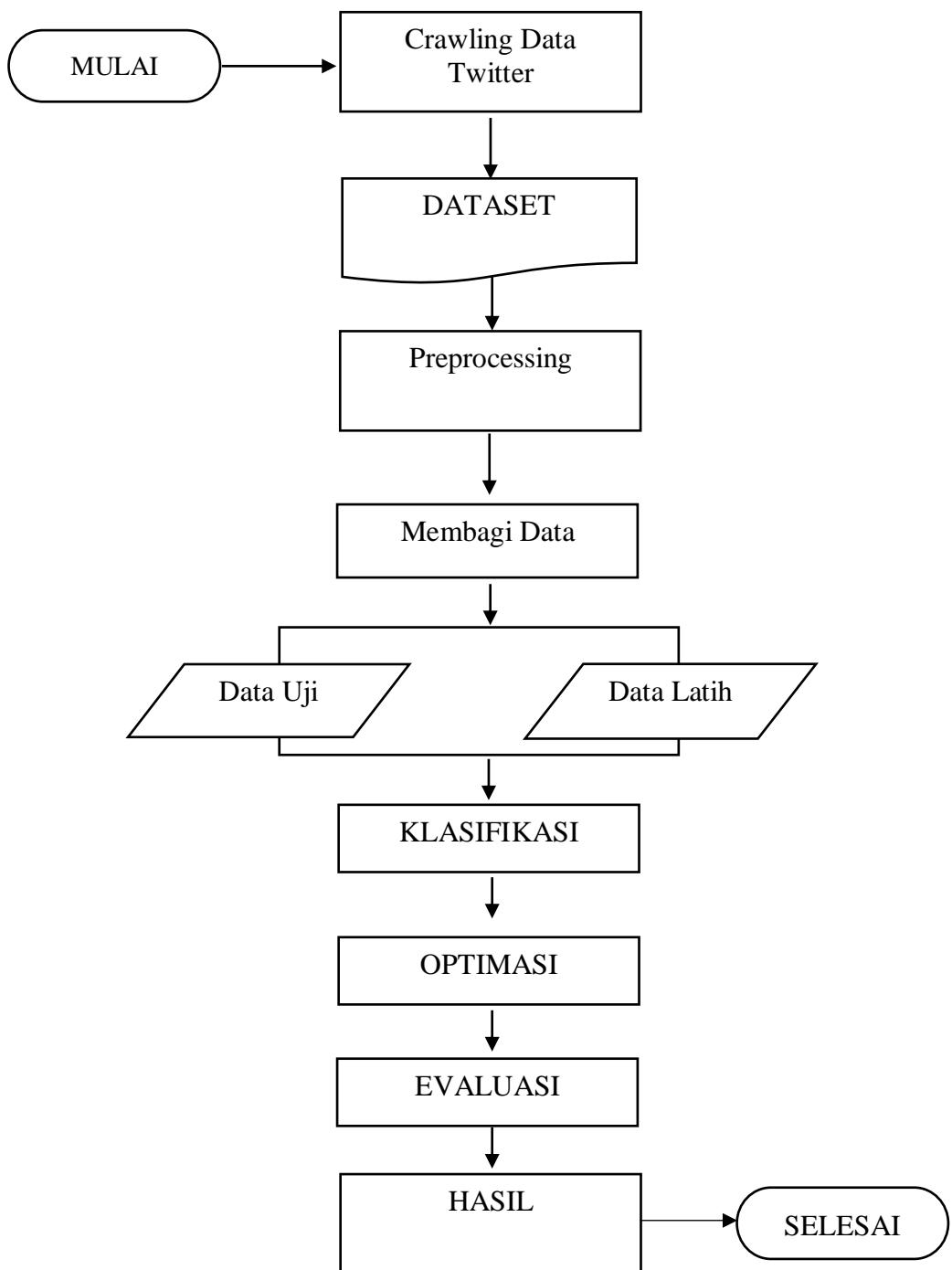
Subject penelitian ini adalah data tweet yang diambil dari platform twitter yang diklasifikasikan menggunakan algoritma Naïve Bayes berbasis PSO dengan object yang diteliti yaitu Ibu Kota Nusantara.

3.1.2 Waktu dan Lokasi

Untuk mendapatkan data-data yang dibutuhkan, maka penulis melakukan penelitian melalui media sosial twitter. Dan penelitian ini akan dimulai dari bulan Mei 2024 sampai Juni 2024.

3.2 Pemodelan Penelitian

Model yang diusulkan dapat digambarkan menggunakan flowchart berikut ini :



Gambar 3. 1 : pemodelan yang diusulkan

3.2.1 Pengumpulan Data

Data tweet yang diperoleh akan diproses pada tahap preprocessing yang terdiri dari :

1. Cleansing

Cleansing adalah proses pembersihan atau penghapusan kata-kata yang dianggap tidak penting. Contoh kata-kata yang dianggap tidak penting meliputi karakter seperti '@', '#', 'rt', dan sejenisnya.

Tabel 3.1 Sampel Cleansing

Teks Input	Teks Output
Ibu Kota Nusantara	Ibu Kota Nusantara
Langkah Strategis untuk	Langkah Strategis untuk
Mengembangkan	Mengembangkan
Indonesia di Masa Depan	Indonesia di Masa Depan
#IKNNusantara	IKNNusantara

2. Case Folding

Case folding adalah proses mengubah semua huruf dalam dokumen tweet menjadi huruf kecil. Hanya huruf dari 'a' hingga 'z' yang akan diproses, sedangkan karakter selain huruf akan diabaikan.

Tabel 3.2 Sampel Case Folding

Teks Input	Teks Output
Ibu Kota Nusantara	ibu kota nusantara langkah
Langkah Strategis untuk	strategis untuk
Mengembangkan	mengembangkan
Indonesia di Masa Depan	indonesia di masa depan
IKNNusantara	iknnusantara

3. Tokenizing

Tokenisasi adalah proses membagi dokumen menjadi bagian-bagian kecil yang bisa berupa bab, sub-bab, paragraf, kalimat, atau kata-kata (token).

Tabel 3.3 Sampel Tokenizing

Teks Input	Teks Output
ibu kota nusantara langkah strategis untuk mengembangkan indonesia di masa depan iknnusantara	Ibu, kota, nusantara, langkah, strategis, untuk, mengembangkan, indonesia, masa, depan, ikn, nusantara

4. Stop Removal

Stopword yaitu proses untuk mengeliminasi kata yang tidak berkaitan dengan kinerja pemerintahan kota gorontalo. Contohnya adalah kata-kata yang umum dan tidak memiliki makna seperti “yang”, “di”, “dalam”, dan lain-lain.

Tabel 3.4 Sampel Sopword Removal

Teks Input	Teks Output
Ibu, kota, nusantara, langkah, strategis, untuk, mengembangkan, indonesia, masa, depan, ikn, nusantara	kota nusantara langkah strategis untuk mengembangkan indonesia masa depan ikn nusantara

5. Stemming

Stemming adalah proses untuk menemukan kata dasar atau 'stem' dari kata-kata yang telah melewati tahap penghapusan stopword (filtering). Dalam

stemming, terdapat dua pendekatan yang umum digunakan, yaitu pendekatan kamus dan pendekatan berdasarkan aturan.

Tabel 3.5 Sampel Stemming

Teks Input	Teks Output
kota nusantara langkah strategis untuk mengembangkan indonesia masa depan ikn nusantara	kota nusantara langkah strategis untuk kembang indonesia masa depan ikn nusantara

6. Normalization

Normalization adalah proses perbaikan kalimat yang memiliki ejaan kata yang disingkat atau diperpanjang menjadi kata baku

Tabel 3.6 Sampel Normalization

Teks Input	Teks Output
kota nusantara langkah strategis untuk kembang indonesia masa depan ikn nusantara	ibu kota nusantara langkah strategis untuk kembang indonesia masa depan ibu kota nusantara nusantara

3.2.2 Proses Klasifikasi

Setelah data selesai diproses melalui tahap preprocessing, langkah berikutnya adalah melakukan klasifikasi data tweet menggunakan metode Naïve Bayes yang akan digunakan dalam penelitian ini. Proses klasifikasi ini bertujuan untuk mengidentifikasi sentimen dari tweet-tweet yang telah terkumpul [16].

3.2.3 Proses Optimasi

Setelah diklasifikasi kemudian dilanjutkan dengan menerapkan fitur seleksi Particle Swarm Optimization (PSO) terhadap algoritma Naive Bayes untuk meningkatkan akurasi pada proses klasifikasi analisis sentimen dibandingkan

dengan tanpa fitur seleksi. Objektif pada penelitian ini adalah adanya peningkatan akurasi pada model Naive Bayes setelah diterapkannya fitur seleksi Particle Swarm Optimization (PSO), di mana pengujian model dilakukan dengan menggunakan teknik Cross Validation dan pengukuran evaluasi menggunakan Confusion Matrix.

3.2.4 Evaluasi

Langkah berikutnya adalah proses evaluasi data yang telah diklasifikasikan dengan menggunakan metode Support Vector Machine, dengan menggunakan Confusion Matrix. Confusion Matrix adalah hasil dari prediksi yang dilakukan dengan metode Support Vector Machine, yang mengukur kinerja tiap kelas dengan menghitung precision, recall, dan f1-score.

3.2.5 Hasil

Menunjukkan hasil dari Penerapan Metode Naïve Bayes yang menggunakan optimasi PSO dengan yang tidak menggunakan Optimasi PSO pada sentiment analisis untuk mengetahui komentar positif, negative terhadap Ibu Kota Nusantara pada twitter.

BAB IV

PEMBANGUNAN MODEL

4.1 Pengumpulan Data

Pada penelitian ini menggunakan dataset *tweet* berbahasa indonesia yang diambil dari akun *twitter*. Kemudian peneliti melakukan pemrosesan dan pengklasifikasian pada dataset berdasarkan pemodelan yang diusulkan. Jika terjadi outlier atau *tweet* yang tidak berhubungan dengan topik yang akan diteliti dan tidak memenuhi kriteria untuk dijadikan dataset pada proses pengumpulan dataset, maka peneliti akan menyeleksi *tweet* tersebut dari dataset agar tidak mengganggu pada saat proses klasifikasi.

Pengumpulan data pada penelitian ini menggunakan *library* pada *python* yang dilakukan dengan menggunakan data scraper dari *tweets* pada media sosial *twitter* yang telah tersedia pada *twitter API (application programming Interface)* dan diambil langsung dari akun media sosial *twitter*. Pengambilan data *tweets* menggunakan *tools scrapping* atau *crawling* data dengan ini peneliti menggunakan *keyword* “Ibu Kota Nusantara”. Berikut adalah tabel hasil *crawling* dari beberapa data yang disimpan.

Tabel 4.1 : Data Crawling

No.	Text
1	@geloraco Ikn jadi kota modern ... Terus kerjakan sampai selesai sesuai amanat uu ikn
2	Presiden terpilih Prabowo Subianto kembali menegaskan sikap dan komitmennya untuk mendukung dan melanjutkan pemindahan ibu kota ke Ibu Kota Nusantara (IKN).
3	Wawancara Dengan Media Timur Tengah Prabowo Subianto Tegaskan Komitmen Untuk Lanjutkan Pembangunan IKN https://t.co/I04Ji1ZPMj
4	@DokterTifa IKN cuma ladang buat bisnis korupsi doang

No.	Text
5	@MafiaWasit @KemenPU Uangnya buat IKN ga sih buat apa menyenangkan rakyat prioritas negara itu pejabat
...	
1364	@geloraco IKN untuk pejabat aja toh

Dari hasil *crawling* data diperoleh 1364 *tweet*. Kemudian dilakukan seleksi pada *tweet* yang memenuhi kriteria untuk diproses sehingga peneliti menyeleksi data *tweet* 1364 menjadi 1116 data *tweet* yang siap untuk diproses.

4.2 Pre-processing

Preprocessing dilakukan setelah dataset siap untuk diolah. Pada tahapan preprocessing terdapat beberapa proses yang akan dilakukan sehingga data menjadi terstruktur dengan tahapan Cleansing, Case folding, Tokenizing, Stopword removal, Stemming, dan Normalization.

Pada tahapan preprocessing peneliti menggunakan *tool google colab* berbasis python, sehingga hasil yang didapatkan dari proses Case folding adalah sebagai berikut :

1) Cleansing

Proses ini dilakukan untuk pembersihan atau penghapusan kata-kata yang dianggap tidak penting. Contoh kata-kata yang dianggap tidak penting meliputi karakter seperti '@', '#', 'rt', dan sejenisnya. Berikut ini adalah sourcecode python pada *google colab* untuk proses cleansing.

```
import re
def cleaningulasan(ulasan):
    ulasan = re.sub('@[^\w]+', '', ulasan)
    ulasan = re.sub('#[^\w]+', '', ulasan)
    ulasan = re.sub('https?://[^\w]+', '', ulasan)
    ulasan = re.sub('[^\w-]+', '', ulasan)
    ulasan = re.sub('([^\w-])\w{2,}<>([^\w-]+)-([^\w-]+\w{2,})', '$1$2$3', ulasan)
    ulasan = ulasan.strip()
    return ulasan
df_data['Cleaning'] = df_data['content'].apply(cleaningulasan)

def cleancemoji(ulasan):
    return ulasan.encode('utf-8', 'ignore').decode('utf-8')
df_data['HapusPunkt'] = df_data['Cleaning'].apply(cleancemoji)

def replaceTOM(ulasan):
    pola = re.compile('([^\w])\w{2,}([^\w])', re.DOTALL)
    return pola.sub('$1$2', ulasan)
df_data['%Lebih'] = df_data['HapusPunkt'].apply(replaceTOM)

df_data_cleaned = df_data.drop_duplicates(subset=['%Lebih'])

df_data_cleaned.to_csv('cleaned_data.csv', index=False)
```

Gambar 4. 1 : Cleaning

dari proses cleansing diatas diperoleh hasil yang lebih bersih dari data sebelumnya . Berikut adalah tabel hasil cleansing dari beberapa data yang disimpan.

Tabel 4.2 : Data Cleansing

No	Text	Cleansing
1	@PartaiSocmed Sama kaya pembuatan IKN pejabatnya ingin buat lingkungan/hunian mewah duitnya dari pajak masyarakat. Sesederhana itu	Sama kaya pembuatan IKN pejabatnya ingin buat lingkungan hunian mewah duitnya dari pajak masyarakat Sesederhana itu
2	@geloraco Apakah IKN di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk ?	Apakah IKN di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk
3	@TOM5helby Bukankah IKN juga produk @DPR_RI ? Makin aneh aja bernegara ini	Bukankah IKN juga produk Makin aneh aja bernegara ini
4	Cocok suruh aja Gibran Nyepi jaga IKN bersama bapaknya dan kroninya disana biar jarang dilihat orang. Kalau di Jkt ntar jd sasaran demo. JKT dan istana negara biar diurus Prabowo jd tdk risi ditunggui si Samsul.	Cocok suruh aja Gibran Nyepi jaga IKN bersama bapaknya dan kroninya disana biar jarang dilihat orang Kalau di Jkt ntar jd sasaran demo JKT dan istana negara biar diurus Prabowo jd tdk risi ditunggui si Samsul
5	@Shanty1BM Negara rugi nangkep trus ngasih makan slama di prodeo. Mending kirim kerja Rodi/Romusha bangun rel di Sumatera Kalimantan Sulawesi atau bikin jln di Papua atau jd kuli buat bangun IKN.	Negara rugi nangkep trus ngasih makan slama di prodeo Mending kirim kerja Rodi Romusha bangun rel di Sumatera Kalimantan Sulawesi atau bikin jln di Papua atau jd kuli buat bangun IKN

No	Text	Cleansing
6	@LexWu_13 @didienAZHAR @jokowi Benar pak. Pelayanan kesehatan masyarakat yg masih kurang diseluruh pelosok tanah air ini sungguh jauh lebih penting daripada proyek IKN atau kereta Whooos.	Benar pak Pelayanan kesehatan masyarakat yg masih kurang diseluruh pelosok tanah air ini sungguh jauh lebih penting daripada proyek IKN atau kereta Whos
7	@toe_giman Amanah undang2 orng miskin dan anak terlantar d pelihara negara ini malah bangun ikn .	Amanah undang orng miskin dan anak terlantar d pelihara negara ini malah bangun ikn
8	@Budiady5 @gibran_tweet aja tugas di IKN daripada ga berguna di ISTANA	aja tugas di IKN daripada ga berguna di ISTANA
9	@DokterTifa Biar IKN yang nempati para oposisi	Biar IKN yang nempati para oposisi
10	Mentan Sebut Sulawesi Barat Bakal Jadi Pemasok Bahan Pangan IKN https://t.co/l8ig3tijFm	Mentan Sebut Sulawesi Barat Bakal Jadi Pemasok Bahan Pangan IKN

2) Case folding

Pada proses ini dilakukan pengubahan huruf besar menjadi huruf kecil pada semua tweet yang ada pada data cleansing. Berikut ini adalah sourcecode python pada *google colab* untuk proses case folding.

```
def casefoldingText(ulasan):
    ulasan = ulasan.lower()
    return ulasan
df_data['CaseFolding']= df_data_cleaned['3/Lebih'].apply(casefoldingText)
```

Gambar 4. 2 : Case Folding

Dari proses case folding diatas diperoleh hasil data yang sudah diubah menjadi huruf kecil. Berikut adalah tabel hasil case folding dari beberapa data yang disimpan.

Tabel 4.3 : Data Case Folding

No	Data Cleaning	Case Folding
1	Sama kaya pembuatan IKN pejabatnya ingin buat lingkungan hunian mewah duitnya dari pajak masyarakat Sesederhana itu	sama kaya pembuatan ikn pejabatnya ingin buat lingkungan hunian mewah duitnya dari pajak masyarakat sesederhana itu
2	Apakah IKN di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk	apakah ikn di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk
3	Benar pak Pelayanan kesehatan masyarakat yg masih kurang diseluruh pelosok tanah air ini sungguh jauh lebih penting daripada proyek IKN atau kereta Whos	benar pak pelayanan kesehatan masyarakat yg masih kurang diseluruh pelosok tanah air ini sungguh jauh lebih penting daripada proyek ikn atau kereta whos
4	Amanah undang orng miskin dan anak terlantar d pelihara negara ini malah bangun ikn	amanah undang orng miskin dan anak terlantar d pelihara negara ini malah bangun ikn
5	Mentan Sebut Sulawesi Barat Bakal Jadi Pemasok Bahan Pangan IKN	mentan sebut sulawesi barat bakal jadi pemasok bahan pangan ikn
6	Gedung dpr segera selesai di ikn yg lama di jkt dijual utk jadi mal	gedung dpr segera selesai di ikn yg lama di jkt dijual utk jadi mal
7	Rezim mementingkan IKN titik	rezim mementingkan ikn titik

No	Data Cleaning	Case Folding
8	Sabar bu uang negara sedang fokus membangun IKN demi kemajuan bangsa dan negara ini	sabar bu uang negara sedang fokus membangun ikn demi kemajuan bangsa dan negara ini
9	Presiden terpilih Prabowo Subianto kembali menegaskan sikap dan komitmennya untuk mendukung dan melanjutkan pemindahan ibu kota ke Ibu Kota Nusantara IKN	presiden terpilih prabowo subianto kembali menegaskan sikap dan komitmennya untuk mendukung dan melanjutkan pemindahan ibu kota ke ibu kota nusantara ikn
10	Pembangunan Ibu Kota Negara IKN Nusantara harus membuat warga Kaltim tambah sejahtera	pembangunan ibu kota negara ikn nusantara harus membuat warga kaltim tambah Sejahtera

3) Tokenizing

Selanjutnya adalah proses pemisahan kalimat (string) yang setiap kalimatnya disusun menjadi potongan tunggal. Berikut ini adalah sourcecode python pada *google colab* untuk proses Tokenisasi

```
nltk.download('punkt')
import pandas as pd
import re
import nltk
from nltk.tokenize import word_tokenize
# Tokenisasi
def tokenizingText(ulasan):
    ulasan = word_tokenize(ulasan)
    return ulasan

df_data['Tokenizing'] = df_data['CaseFolding'].apply(tokenizingText)

# Tampilkan hasil
df_data[['CaseFolding', 'Tokenizing']]
```

Gambar 4. 3 : Tokenizing

Dari proses tokenizing diperoleh hasil data yang telah dipisahkan dari kalimat. Berikut adalah tabel hasil tokenizing dari beberapa data yang disimpan.

Tabel 4.4 : Data Tokenizing

No	Case Folding	Tokenizing
1	sama kaya pembuatan ikn pejabatnya ingin buat lingkungan hunian mewah duitnya dari pajak masyarakat sesederhana itu	['sama', 'kaya', 'pembuatan', 'ikn', 'pejabatnya', 'ingin', 'buat', 'lingkungan', ' hunian', 'mewah', 'duitnya', 'dari', 'pajak', 'masyarakat', 'sesederhana', 'itu']
2	apakah ikn di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk	['apakah', 'ikn', 'di', 'bangun', 'untuk', 'menjadi', 'kota', 'tertutup', 'dan', 'hanya', 'orang', 'tertentu', 'saja', 'yg', 'boleh', 'masuk']
3	benar pak pelayanan kesehatan masyarakat yg masih kurang diseluruh pelosok tanah air ini sungguh jauh lebih penting daripada proyek ikn atau kereta whos	['benar', 'pak', 'pelayanan', 'kesehatan', 'masyarakat', 'yg', 'masih', 'kurang', 'diseluruh', 'pelosok', 'tanah', 'air', 'ini', 'sungguh', 'jauh', 'lebih', 'penting', 'daripada', 'proyek', 'ikn', 'atau', 'kereta', 'whos']
4	amanah undang orng miskin dan anak terlantar d pelihara negara ini malah bangun ikn	['amanah', 'undang', 'orng', 'miskin', 'dan', 'anak', 'terlantar', 'd', 'pelihara', 'negara', 'ini', 'malah', 'bangun', 'ikn']
5	mentan sebut sulawesi barat bakal jadi pemasok bahan pangan ikn	['mentan', 'sebut', 'sulawesi', 'barat', 'bakal', 'jadi', 'pemasok', 'bahan', 'pangan', 'ikn']
6	gedung dpr segera selesai di ikn yg lama di jkt dijual utk jadi mal	['gedung', 'dpr', 'segera', 'selesai', 'di', 'ikn', 'yg', 'lama', 'di', 'jkt', 'dijual', 'utk', 'jadi', 'mal']
7	rezim mementingkan ikn titik	['rezim', 'mementingkan', 'ikn', 'titik']

No	Case Folding	Tokenizing
8	sabar bu uang negara sedang fokus membangun ikn demi kemajuan bangsa dan negara ini	['sabar', 'bu', 'uang', 'negara', 'sedang', 'fokus', 'membangun', 'ikn', 'demi', 'kemajuan', 'bangsa', 'dan', 'negara', 'ini']
9	presiden terpilih prabowo subianto kembali menegaskan sikap dan komitmennya untuk mendukung dan melanjutkan pemindahan ibu kota ke ibu kota nusantara ikn	['presiden', 'terpilih', 'prabowo', 'subianto', 'kembali', 'menegaskan', 'sikap', 'dan', 'komitmennya', 'untuk', 'mendukung', 'dan', 'melanjutkan', 'pemindahan', 'ibu', 'kota', 'ke', 'ibu', 'kota', 'nusantara', 'ikn']
10	pembangunan ibu kota negara ikn nusantaraharus membuat warga kaltim tambah Sejahtera	['pembangunan', 'ibu', 'kota', 'negara', 'ikn', 'nusantaraharus', 'membuat', 'warga', 'kaltim', 'tambah', 'sejahtera']

4) Stopword Removal

Pada proses ini terjadi penghilangan kata yang tidak berkaitan dengan topik dokumen. Berikut ini adalah sourcecode python pada *google colab* untuk proses Stopword removal

Gambar 4. 4 : Stopword Removal

Dari proses stopword diatas diperoleh hasil data yang telah dihilangkan karena tidak berkaitan dengan topik. Berikut adalah tabel hasil stopword dari beberapa data yang disimpan.

Tabel 4.5 : Data Stopword Removal

No	Tokenizing	Stopword Removal
1	['sama', 'kaya', 'pembuatan', 'ikn', 'pejabatnya', 'ingin', 'buat', 'lingkungan', 'hunian', 'mewah', 'duitnya', 'dari', 'pajak', 'masyarakat', 'sesederhana', 'itu']	['kaya', 'pembuatan', 'ikn', 'pejabatnya', 'lingkungan', 'hunian', 'mewah', 'duitnya', 'pajak', 'masyarakat', 'sesederhana']
2	['apakah', 'ikn', 'di', 'bangun', 'untuk', 'menjadi', 'kota', 'tertutup', 'dan', 'hanya', 'orang', 'tententu', 'saja', 'yg', 'boleh', 'masuk']	['ikn', 'bangun', 'kota', 'tertutup', 'orang', 'masuk']
3	['benar', 'pak', 'pelayanan', 'kesehatan', 'masyarakat', 'yg', 'masih', 'kurang', 'diseluruh', 'pelosok', 'tanah', 'air', 'ini', 'sungguh', 'jauh', 'lebih', 'penting', 'daripada', 'proyek', 'ikn', 'atau', 'kereta', 'whos']	['pelayanan', 'kesehatan', 'masyarakat', 'diseluruh', 'pelosok', 'tanah', 'air', 'sungguh', 'proyek', 'ikn', 'kereta', 'whos']
4	['amanah', 'undang', 'orng', 'miskin', 'dan', 'anak', 'terlantar', 'd', 'pelihara', 'negara', 'ini', 'malah', 'bangun', 'ikn']	['amanah', 'undang', 'orng', 'miskin', 'anak', 'terlantar', 'pelihara', 'negara', 'bangun', 'ikn']
5	['mentan', 'sebut', 'sulawesi', 'barat', 'bakal', 'jadi', 'pemasok', 'bahan', 'pangan', 'ikn']	['mentan', 'sulawesi', 'barat', 'pemasok', 'bahan', 'pangan', 'ikn']
6	['gedung', 'dpr', 'segera', 'selesai', 'di', 'ikn', 'yg', 'lama', 'di', 'jkt', 'dijual', 'utk', 'jadi', 'mal']	['gedung', 'dpr', 'selesai', 'ikn', 'jkt', 'dijual', 'mal']

No	Tokenizing	Stopword Removal
7	['rezim', 'mementingkan', 'ikn', 'titik']	['rezim', 'mementingkan', 'ikn', 'titik']
8	['sabar', 'bu', 'uang', 'negara', 'sedang', 'fokus', 'membangun', 'ikn', 'demi', 'kemajuan', 'bangsa', 'dan', 'negara', 'ini']	['sabar', 'bu', 'uang', 'negara', 'fokus', 'membangun', 'ikn', 'kemajuan', 'bangsa', 'negara']
9	['presiden', 'terpilih', 'prabowo', 'subianto', 'kembali', 'menegaskan', 'sikap', 'dan', 'komitmennya', 'untuk', 'mendukung', 'dan', 'melanjutkan', 'pemindahan', 'ibu', 'kota', 'ke', 'ibu', 'kota', 'nusantara', 'ikn']	['presiden', 'terpilih', 'prabowo', 'subianto', 'sikap', 'komitmennya', 'mendukung', 'melanjutkan', 'pemindahan', 'kota', 'kota', 'nusantara', 'ikn']
10	['pembangunan', 'ibu', 'kota', 'negara', 'ikn', 'nusantaraharus', 'membuat', 'warga', 'kaltim', 'tambah', 'sejahtera']	['pembangunan', 'kota', 'negara', 'ikn', 'nusantaraharus', 'warga', 'kaltim', 'sejahtera']

5) Steamming

Proses ini dilakukan penghapusan imbuhan dari setiap kata. Sehingga mendapatkan kata dasarnya. Berikut ini adalah sourcecode python pada *google colab* untuk proses Steamming.

```

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import os

factory = StemmerFactory()
steamer = factory.create_stemmer()

def stemmed_wrapper(term):
    return steamer.stem(term)

term_dict = {}

for document in df_data['Stopword Removal']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ""

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":", term_dict[term])

def stemming_text(document):
    return [term_dict[term] for term in document]

df_data['Stemming'] = df_data['Stopword Removal'].apply(stemming_text)
df_data

```

Gambar 4. 5 : Steamming

Setelah proses steamming diatas diperoleh hasil data yang berbentuk kata dasar. Berikut adalah tabel hasil steamming dari beberapa data yang disimpan.

Tabel 4.6 : Data Steamming

No	Stopword Removal	Steamming
1	['kaya', 'pembuatan', 'ikn', 'pejabatnya', 'lingkungan', 'hunian', 'mewah', 'duitnya', 'pajak', 'masyarakat', 'sesederhana']	['kaya', 'buat', 'ikn', 'jabat', 'lingkung', 'huni', 'mewah', 'duit', 'pajak', 'masyarakat', 'sederhana']
2	['ikn', 'bangun', 'kota', 'tertutup', 'orang', 'masuk']	['ikn', 'bangun', 'kota', 'tutup', 'orang', 'masuk']
3	['pelayanan', 'kesehatan', 'masyarakat', 'diseluruh', 'pelosok', 'tanah', 'air', 'sungguh', 'proyek', 'ikn', 'kereta', 'whos']	['layan', 'sehat', 'masyarakat', 'seluruh', 'pelosok', 'tanah', 'air', 'sungguh', 'proyek', 'ikn', 'kereta', 'whos']
4	['amanah', 'undang', 'orng', 'miskin', 'anak', 'terlantar', 'pelihara', 'negara', 'bangun', 'ikn']	['amanah', 'undang', 'orng', 'miskin', 'anak', 'lantar', 'pelihara', 'negara', 'bangun', 'ikn']
5	['mentan', 'sulawesi', 'barat', 'pemasok', 'bahan', 'pangan', 'ikn']	['tan', 'sulawesi', 'barat', 'pasok', 'bahan', 'pangan', 'ikn']
6	['gedung', 'dpr', 'selesai', 'ikn', 'jkt', 'dijual', 'mal']	['gedung', 'dpr', 'selesai', 'ikn', 'jkt', 'jual', 'mal']
7	['rezim', 'mementingkan', 'ikn', 'titik']	['rezim', 'penting', 'ikn', 'titik']

No	Stopword Removal	Steamming
8	['sabar', 'bu', 'uang', 'negara', 'fokus', 'membangun', 'ikn', 'kemajuan', 'bangsa', 'negara']	['sabar', 'bu', 'uang', 'negara', 'fokus', 'bangun', 'ikn', 'maju', 'bangsa', 'negara']
9	['presiden', 'terpilih', 'prabowo', 'subianto', 'sikap', 'komitmennya', 'mendukung', 'melanjutkan', 'pemindahan', 'kota', 'kota', 'nusantara', 'ikn']	['presiden', 'pilih', 'prabowo', 'subianto', 'sikap', 'komitmen', 'dukung', 'lanjut', 'pindah', 'kota', 'kota', 'nusantara', 'ikn']
10	['pembangunan', 'kota', 'negara', 'ikn', 'nusantaraharus', 'warga', 'kaltim', 'sejahtera']	['bangun', 'kota', 'negara', 'ikn', 'nusantaraharus', 'warga', 'kaltim', 'sejahtera']

6) Normalization

Pada proses ini kalimat yang memiliki ejaan kata yang disingkat atau diperpanjang akan diperbaiki menjadi kata baku. Berikut ini adalah sourcecode python pada *google colab* untuk proses normalization.

```
def convertToSlangword(ulasan):
    kamusSlang = eval(open("slangwords.txt").read())
    pattern = re.compile(r'\b(' + '|'.join(kamusSlang.keys())+r')\b')
    content = []
    for kata in ulasan:
        filterSlang = pattern.sub(lambda x: kamusSlang[x.group()],kata)
        content.append(filterSlang.lower())
    ulasan = content
    return ulasan

df_data['Normalization'] = df_data['Stemming'].apply(convertToSlangword)

df_data
```

Gambar 4. 6 : Normalization

Dari proses normalization menghasilkan data yang telah diperbaiki menjadi lebih baku. Berikut adalah tabel hasil normalization dari beberapa data yang disimpan.

Tabel 4.7 : Data Normalization

No.	Steamming	Normalization
1	['kaya', 'buat', 'ikn', 'jabat', 'lingkung', 'huni', 'mewah', 'duit', 'pajak', 'masyarakat', 'sederhana']	kaya buat ikn jabat lingkung huni mewah uang pajak masyarakat sederhana
2	['ikn', 'bangun', 'kota', 'tutup', 'orang', 'masuk']	ikn bangun kota tutup orang masuk
3	['layan', 'sehat', 'masyarakat', 'seluruh', 'pelosok', 'tanah', 'air', 'sungguh', 'proyek', 'ikn', 'kereta', 'whos']	layan sehat masyarakat seluruh pelosok tanah air sungguh proyek ikn kereta whos
4	['amanah', 'undang', 'orng', 'miskin', 'anak', 'lantar', 'pelihara', 'negara', 'bangun', 'ikn']	amanah undang orang miskin anak lantar pelihara negara bangun ikn
5	['sulawesi', 'barat', 'pasok', 'bahan', 'pangan', 'ikn']	sulawesi barat pasok bahan pangan ikn
6	['gedung', 'dpr', 'selesai', 'ikn', 'jkt', 'jual', 'mal']	gedung dpr selesai ikn jakarta jual mal
7	['rezim', 'penting', 'ikn', 'titik']	rezim penting ikn titik
8	['sabar', 'uang', 'negara', 'fokus', 'bangun', 'ikn', 'maju', 'bangsa', 'negara']	Sabar uang negara fokus bangun ikn maju bangsa negara

No.	Steamming	Normalization
9	['presiden', 'pilih', 'prabowo', 'subianto', 'sikap', 'komitmen', 'dukung', 'lanjut', 'pindah', 'kota', 'kota', 'nusantara', 'ikn']	presiden pilih prabowo subianto sikap komitmen dukung lanjut pindah kota kota nusantara ikn
10	['bangun', 'kota', 'negara', 'ikn', 'nusantaraharus', 'warga', 'kaltim', 'sejahtera']	bangun kota negara ikn nusantaraharus warga kaltim sejahtera

4.3 Pelabelan Data

Tahapan selanjutnya setelah preprocessing dilakukan, kemudian peneliti memberikan pelabelan dengan menggunakan *library textblob* secara otomatis. Namun sebelum melakukan pelabelan, teks harus diterjemahkan terlebih dahulu kedalam bahasa inggris mengingat *textblob* hanya mampu melabeli teks berbahasa inggris saja. Pada tahapan pelabelan peneliti menggunakan *google colab* berbasis python, sehingga hasil yang didapatkan dari proses *translation* adalah sebagai berikut.

1.) Translation

Pada tahapan ini peneliti menerjemahkan kalimat dari bahasa Indonesia kedalam bahasa Inggris dengan bantuan *google translate*. Dataset yang digunakan adalah dataset yang telah melewati proses pre-processing mulai dari cleansing, case folding, tokenizing, stop removal, stemming, dan normalization. Berikut ini adalah sourcecode python pada *google colab* untuk proses translation.

```

from googletrans import Translator
# Membuat instance dari Translator
translator = Translator()

# Fungsi untuk melakukan translate teks
def translate_text(ulasan, src='id', dest='en'):
    try:
        # Melakukan translate
        translation = translator.translate(ulasan, src=src, dest=dest)
        return translation.text
    except Exception as e:
        # Jika terjadi error, kembalikan teks asli
        print(f'Error during translation: {e}')
        return ulasan

# Terapkan fungsi translate pada kolom 'Normalization String'
df_data['Translated_Text'] = df_data['Caselolding'].apply(lambda text: translate_text(text))

# Menampilkan Dataframe untuk memverifikasi hasil translate
print(df_data['Caselolding', 'Translated_Text'])
df_data.to_csv('translated_data.csv', index=False)

df_data

```

Gambar 4.7 : Translation

Dari proses tranlation menghasilkan data yang telah diterjemahkan kedalam bahasa inggris. Berikut adalah tabel hasil translation dari beberapa data yang disimpan.

Tabel 4.8 : Data Translation

No.	Text	Translation
1	sama kaya pembuatan ikn pejabatnya ingin buat lingkungan hunian mewah duitnya dari pajak masyarakat sesederhana itu	as rich as the manufacture of officials, its officials want to make a luxurious money residential environment from public taxes that simple
2	apakah ikn di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk	Is the IKN built to be a closed city and only certain people can enter
3	pembangunan ikn menurut prabowo subianto akan tetap melindungi serta menjamin kesejahteraan masyarakat lokal dan masyarakat adat yang berada di ikn	The development of IKN according to Prabowo Subianto will continue to protect and guarantee the welfare of the local community and indigenous peoples who are in the IKN

No.	Text	Translation
4	dalam wawancara tersebut prabowo juga menegaskan komitmennya untuk melindungi serta menjamin kesejahteraan masyarakat adat yang berada di ikn	In the interview Prabowo also emphasized his commitment to protect and guarantee the welfare of indigenous peoples who were in the IKN
5	i kn cuma ladang buat bisnis korupsi doang	IKN is only a field for the corruption business
6	mari kita dukung pembangunan i kn	Let's support the development of the IKN
7	prabowo tegaskan komitmen untuk kesejahteraan masyarakat adat di i kn	Prabowo emphasizes the commitment for the welfare of indigenous peoples in IKN
8	i kn ini emang dirancang sebagai kota yang maju dan bakalan sering banget pake inovasi baru yang belum ada di kota kota sebelumnya terutama di bidang teknologi	IKN is indeed designed as a developed city and will often use new innovations that do not yet exist in the previous city, especially in the field of technology
9	i kn nusantara akan jadi hunian untuk generasi indonesia di masa depan	IKN Nusantara will be a residence for the Indonesian generation in the future
10	i kn bawa dampak baik bagi indonesia	IKN Bring a good impact on Indonesia

2.) Pelabelan Data

Pada penelitian ini data tweet yang digunakan hanya dua kategori tweet, yaitu tweet positif dan tweet negatif. Dengan menggunakan *library* *textblob* menghasilkan data sentimen dengan kategori *tweet* positif. Berikut ini adalah sourcecode python pada *google colab* untuk proses pelabelan

```
from textblob import TextBlob

def sentiment_analysis(text):
    blob = TextBlob(text)
    polarity = blob.sentiment.polarity
    if polarity > 0:
        return 'Positive'
    else:
        return 'Negative'

# Contoh penggunaan pada DataFrame setelah translate
# Anda dapat mengganti 'Translated_Text' dengan nama kolom yang sesuai
df_data['Sentiment_Label'] = df_data['Translated_Text'].apply(sentiment_analysis)

# Menampilkan DataFrame untuk memeriksa hasil pelabelan
df_data[['Translated_Text', 'Sentiment_Label']]

# Menyimpan DataFrame ke dalam file CSV jika perlu
df_data.to_csv('Labeled_Data.csv', index=False)

# Menghitung jumlah positif dan negatif
positif_count = (df_data['Sentiment_Label'] == 'Positive').sum()
negatif_count = (df_data['Sentiment_Label'] == 'Negative').sum()

print("Jumlah Sentimen Positif:", positif_count)
print("Jumlah Sentimen Negatif:", negatif_count)
```

Gambar 4.8 : Pelabelan

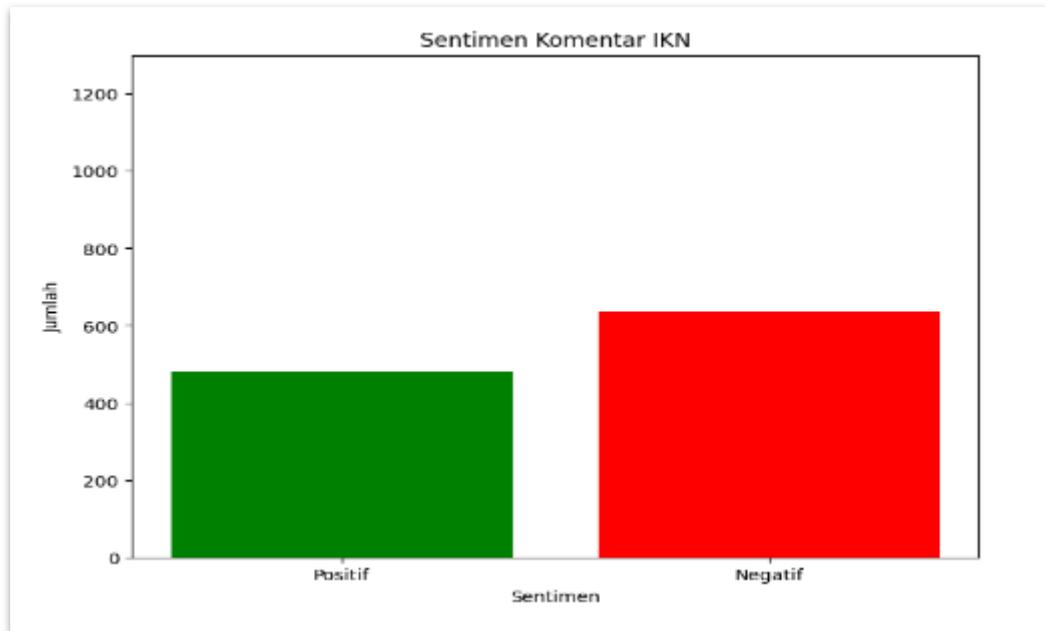
Dari proses pelabelan diatas diperoleh hasil data sentimen positif sebanyak 636 tweet dan sentimen negatif sebanyak 480 tweet. Berikut adalah tabel hasil translation dari beberapa data yang disimpan.

Tabel 4.9 : Data Pelabelan

No.	Text	Label
1	ikn jadi kota modern terus kerjakan sampai selesai sesuai amanat uu ikn	Positif
2	rezim mementingkan ikn titik	Negatif
3	bukan hanya untuk pengembangan wilayah kalimantan timur tempat ikn berada pemindahan ibu kota juga penting untuk keberlangsungan jakarta	Positif
4	ikn cuma ladang buat bisnis korupsi doang	Negatif

No.	Text	Label
5	pemerintah terus berupaya membangun sistem pertahanan cerdas atau smart defense system di ikn nusantara sistem pertahanan cerdas harus dibangun di ikn karena daerah tersebut nantinya akan menjadi pusat pemerintahan indonesia	Positif
6	pemerintah terus berupaya membangun sistem pertahanan cerdas di ikn nusantara sistem pertahanan cerdas harus dibangun di ikn karena daerah tersebut nantinya akan menjadi pusat pemerintahan indonesia serta sesuai dengan rencana induk sistem pertahanan negara di ikn	Positif
7	zero stunting bukan sekadar mimpi di ikn tapi komitmen nyata untuk menciptakan generasi penerus yang sehat dan Tangguh	Positif
8	jual saja semuanya pada asing biar ikn jadi sewa sama asing ambisi sih boleh jika ngutang terus ya bangkrut woy	Negatif
9	ikn ini emang dirancang sebagai kota yang maju dan bakalan sering banget pake inovasi baru yang belum ada di kota kota sebelumnya terutama di bidang teknologi	Positif
10	ikn bawa dampak baik bagi indonesia	Positif

Berikut visualisasi dari kategori *tweet* positif dan negatif



Gambar 4.9 : Visualisasi Sentimen

4.4 Hasil Pembobotan TF IDF

Tahapan selanjutnya setelah preprocessing dan pelabelan dilakukan, kemudian peneliti memberikan bobot pada setiap kata dengan menggunakan pembobotan TF-IDF. Pada tahap ini dilakukan pembobotan kata dari hasil stemming dengan metode Term Inverse Document Frequency (TF-IDF). Pembobotan kata bertujuan untuk menentukan seberapa penting setiap kata dalam dokumen atau baris data. Tiap kata memiliki bobot berbeda karena dipengaruhi oleh frekuensi kemunculan kata tersebut pada tiap dokumen dan banyaknya dokumen yang terdapat kata tersebut.

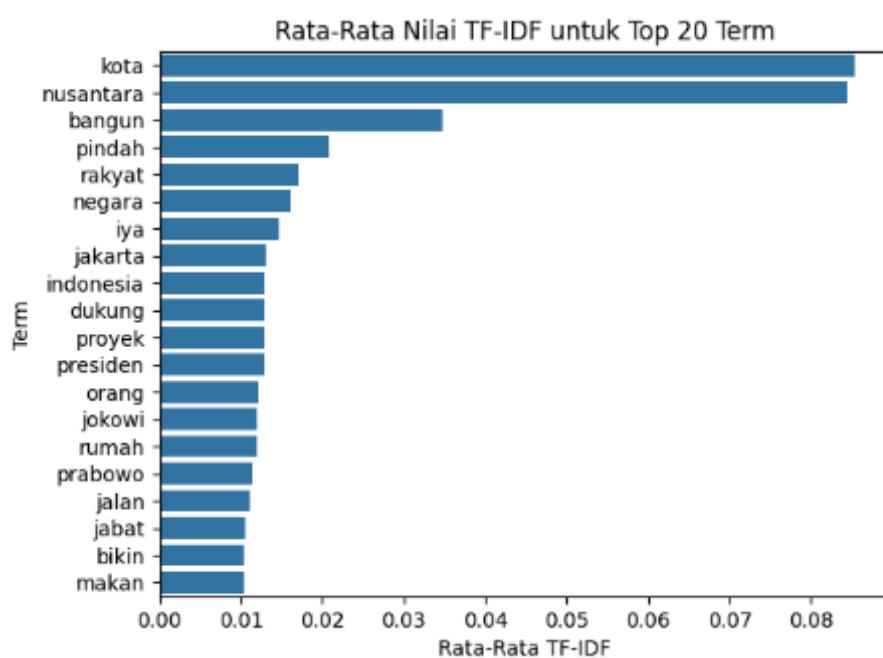
Library yang dibutuhkan dalam tahap ini adalah Scikit-Learn, yang mana library ini merupakan library yang sangat terkenal yang digunakan pada pemodelan data, machine learning dan juga deep learning. Scikit-Learn berisi serangkaian modul atau fungsi yang dapat dipanggil untuk digunakan. Adapun modul atau fungsi yang digunakan pada tahap ini adalah TfIdfVectorizer, dimana modul ini

akan mengonversi kumpulan dokumen data ulasan menjadi matriks fitur TF-IDF. Berikut script Python dari proses ekstraksi fitur menggunakan library

```
# Vektorisasi Teks
vectorizer = TfidfVectorizer(max_features=5000)
X = vectorizer.fit_transform(df_data['Stemming'])
y = df_data['Sentiment_Label'].apply(lambda x: 1 if x == 'Positive' else 0)
```

Gambar 4.10 : TF-IDF

Setelah data melewati proses TF-IDF ini maka tiap kata pada setiap baris data akan memiliki nilai bobot. Nilai bobot tentunya dipengaruhi oleh seberapa sering kata atau seberapa banyak frekuensi data itu muncul dalam dataset yang kita miliki. Adapun beberapa hasil rata-rata pembobotan dengan menggunakan metode TF-IDF ditampilkan pada gambar dibawah ini.



Gambar 4.11 : Visualisasi Top 20 Term

Adapun hasil dari TF-IDF ini juga dapat kita visualisasikan dalam bentuk wordcloud, yang mana semakin besar bobot TF-IDF nya maka kata itu akan menentukan ukuran dari kata yang akan diplot pada wordcloud.



Gambar 4.12 : Word Cloud TF-IDF

Dari visualisasi wordcloud dapat dilihat bahwa kata-kata yang divisualisasikan dengan ukuran besar memiliki frekuensi kemunculan yang tinggi.

4.5 Klasifikasi

Pada tahapan ini akan dilakukan proses klasifikasi dengan menggunakan metode algoritma Naïve Bayes dengan penambahan seleksi fitur menggunakan Particle Swarm Optimization (PSO). Skenario pada tahapan klasifikasi telah ditentukan, dengan melakukan pengujian sebanyak 2 kali. Pada pengujian pertama, dilakukan pengujian menggunakan Naïve Bayes yang tidak dilakukan penambahan seleksi fitur dengan Particle Swarm Optimization (PSO). Dan untuk pada pengujian kedua akan dilakukan menggunakan Naïve Bayes dan terdapat penambahan seleksi fitur dengan Particle Swarm Optimization (PSO). Proses Klasifikasi Naïve Bayes dilakukan Setelah hasil Dokumen-dokumen tersebut diberikan bobot menggunakan TF-IDF. Pada tahapan ini peneliti melakukan klasifikasi menggunakan algoritma Naïve Bayes classifier dengan library Scikit-Learn.

Selanjutnya sistem dapat melakukan pengujian menggunakan data uji dengan tujuan menguji ketepatan suatu sistem dalam mengklasifikasikan data.

4.6 Pembagian Data training dan Testing

Sebelum tahap klasifikasi dilakukan terlebih dahulu akan dilakukan pembagian dataset menjadi dua yaitu data latih dan data uji. Data latih akan digunakan untuk melatih model sedangkan data uji akan di gunakan untuk menguji model yang telah dilatih agar dapat diketahui seberapa baik model yang telah dibuat. Pembagian data untuk data training dan data testing menggunakan perbandingan 80% : 20%.

Pembagian dataset ini dapat dilakukan dengan menggunakan modul atau fungsi dari library python yaitu modul `train_test_split` dari library Scikit-Learn. Adapun script python dari pembagian dataset ini adalah sebagai berikut :

```
# Memisahkan data menjadi training dan testing
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Gambar 4.13 : Pembagian data latih dan data tes

Setelah pembagian dataset dilakukan maka data siap diterapkan pada model. Model pertama *Naïve Bayes* tanpa Particle Swarm Optimization (PSO) dibuat dengan menggunakan modul `naïve_bayes` dari library *Scikit- Learn* Python. Parameter model *Naïve Bayes* yang akan digunakan pada model klasifikasi ini adalah *multinomial naïve bayes*. Pada text classification dari ketiga model Naïve Bayes yang paling sesuai adalah *Multinomial Naïve Bayes* karena jenis data kita berupa diskrit, sedangkan dua lainnya adalah untuk jenis data *binary* dan *continuous*. Berikut script python dari model *Naïve Bayes* :

```

# Latih dan evaluasi model Naive Bayes tanpa optimasi
model_no_opt = MultinomialNB()
model_no_opt.fit(X_train, y_train)
y_pred_no_opt = model_no_opt.predict(X_test)
y_prob_no_opt = model_no_opt.predict_proba(X_test)

# Evaluasi kinerja model tanpa optimasi
accuracy_no_opt = accuracy_score(y_test, y_pred_no_opt)
fpr_no_opt, tpr_no_opt, _ = roc_curve(y_test, y_prob_no_opt[:, 1], pos_label=model_no_opt.classes_[1])
roc_auc_no_opt = auc(fpr_no_opt, tpr_no_opt)

print(f'Non-optimized Accuracy: {accuracy_no_opt}')
print(f'Non-optimized AUC: {roc_auc_no_opt}')

```

Gambar 4.14 : Naive Bayes tanpa PSO

Kemudian peneliti melakukan perhitungan untuk mendapatkan hasil klasifikasi berdasarkan model yang di bangun. Pada perhitungan ini total data training menggunakan 10 dokumen dengan jumlah sentimen positif ada 6 dokumen dan sentimen negatif ada 4 dokumen. Berikut adalah tabel dari data training.

Tabel 4.10 : Data *Training*

Doc	Terms	Label
D1	kota nusantara jadi kota modern terus kerjakan sampai selesai sesuai amanat kota nusantara	Positive
D2	kota nusantara cuma ladang buat bisnis korupsi doang	Negative
D3	kota nusantara nusantara akan memperkuat ketahanan ekonomi dan mengurangi ketergantungan impor baja	Positive
D4	mantap kota nusantara	Positive
D5	kota nusantara ajang lahan basah cuma mau cari cuan saja akal bulus mereka coba buka daftar harta para pejabat yang terlibat kota nusantara kita lihat apa benar bersih hartanya	Negative
D6	kota nusantara hebat	Positive
D7	kota nusantara adalah hambalang jilid tapi lebih parah	Negative
D8	kelihatan kotornya pikiran mereka cuma mau ngeruk keuntungan dari proyek kota nusantara yang akan dinikmati turun menurun sampai dunia kiamat	Negative

Doc	Terms	Label
D9	kota nusantara lebih mengerikan	Negative
D10	kota nusantara proyek tidak jelas	Negative

Selanjutnya untuk Dokumen D11 yang disajikan pada Tabel dibawah ini adalah contoh dari data testing.

Tabel 4.11 : Data Testing

D11	kota nusantara cerdas	?
-----	-----------------------	---

Selanjutnya adalah perhitungan probabilitas setiap kategori kelas dari data training, yaitu probabilitas sentimen positif dan sentimen negatif. Berikut adalah perhitungan probabilitas setiap kategori kelas.

$$P(\text{Positif}) = \frac{4}{10} = 0,4$$

$$P(\text{Negatif}) = \frac{6}{10} = 0,6$$

Setelah menghitung probabilitas setiap kategori kelas, setiap kata atau term yang muncul dalam dokumen D11 dihitung probabilitasnya. Tabel dibawah ini menunjukkan perhitungan probabilitas setiap kata yang muncul dalam dokumen D11

Tabel 4.12 : Hasil hitung probabilitas kata Doc 11

Probabilitas Kata Untuk Sentimen Positif	Probabilitas Kata Untuk Sentimen Negatif
$P(\text{"kota"} \text{Pos} = \frac{0,279 + 1}{32 + 83} = 0,0111)$	$P(\text{"kota"} \text{Neg} = \frac{0,169 + 1}{70 + 83} = 0,00764)$

Probabilitas Kata Untuk sentimen Positif	Probabilitas Kata Untuk Sentimen Negatif
$P("nusantara" Pos = \frac{0,281 + 1}{32 + 83} = 0,01114)$	$P("nusantara" Neg = \frac{0,169 + 1}{70 + 83} = 0,00764)$

Selanjutnya adalah menghitung probabilitas total dari data testing, yaitu probabilitas dari dokumen 11. Probabilitasnya dihitung untuk setiap kategori kelas, yaitu sentimen positif dan sentimen negatif. Berikut perhitungan dari probabilitas total data testing untuk setiap kategori kelas.

$$P(D11 | Pos) = 0,01112 \times 0,01114 \times 0,4 = 0,00004955072$$

$$P(D11 | Neg) = 0,00764 \times 0,00764 \times 0,6 = 0,00003502176$$

Dari perhitungan di atas, nilai probabilitas sentimen positif adalah 0,00004955072 sedangkan nilai probabilitas sentimen negatifnya adalah 0,00003502176. Nilai probabilitas sentimen Positif dokumen D11 lebih besar dari pada nilai probabilitas Negatif sehingga dokumen D11 diklasifikasikan ke dalam sentimen Positif.

Selanjutnya Model kedua adalah dengan menggunakan model Naïve Bayes dan menambahkan seleksi fitur dari *Particle Swarm Optimization* (PSO). Hasil dari model ini adalah untuk mengetahui tingkat keakuratan suatu model klasifikasi yang berbasis seleksi fitur. Berikut script python dari model Naïve Bayes dengan *Particle Swarm Optimization* (PSO)

```

# Fungsi objektif yang mengambilkan negatif dari akurasi Naive Bayes
def nb_accuracy(alpha):
    model = MultinomialNB(alpha=alpha[0])
    scores = cross_val_score(model, X_train, y_train, cv=10, scoring='accuracy')
    return -np.mean(scores) # PSO menimilkan fungsi, jadi kita balik tanda

# Batas atas dan bawah untuk alpha
lb = [0.01]
ub = [10]

# Optimasi menggunakan PSO
alpha_opt, fopt = pso(nb_accuracy, lb, ub, swarmsize=10, omega=1.0, phip=1.0, phig=1.0, maxiter=100, minstep=1e-8)
print(f'Best alpha: {alpha_opt}')

# Latih model dengan alpha optimal dan mengevaluasi kinerja
model_opt = MultinomialNB(alpha=alpha_opt[0])
model_opt.fit(X_train, y_train)
y_pred_opt = model_opt.predict(X_test)
y_prob_opt = model_opt.predict_proba(X_test)

# Evaluasi kinerja model dengan alpha optimal
accuracy_opt = accuracy_score(y_test, y_pred_opt)
fpr_opt, tpr_opt, _ = roc_curve(y_test, y_prob_opt[:, 1], pos_label=model_opt.classes_[1])
roc_auc_opt = auc(fpr_opt, tpr_opt)

print(f'Optimal Accuracy: {accuracy_opt}')
print(f'Optimal AUC: {roc_auc_opt}')

```

Gambar 4.15 : Naive Bayes dengan PSO

Dapat dilihat bahwa dari script kita menguji model naive bayes menggunakan *Particle Swarm Optimization* (PSO). Dari proses ini *Particle Swarm Optimization* (PSO) mencari nilai optimal guna meningkatkan nilai akurasi dari model Naive Bayes

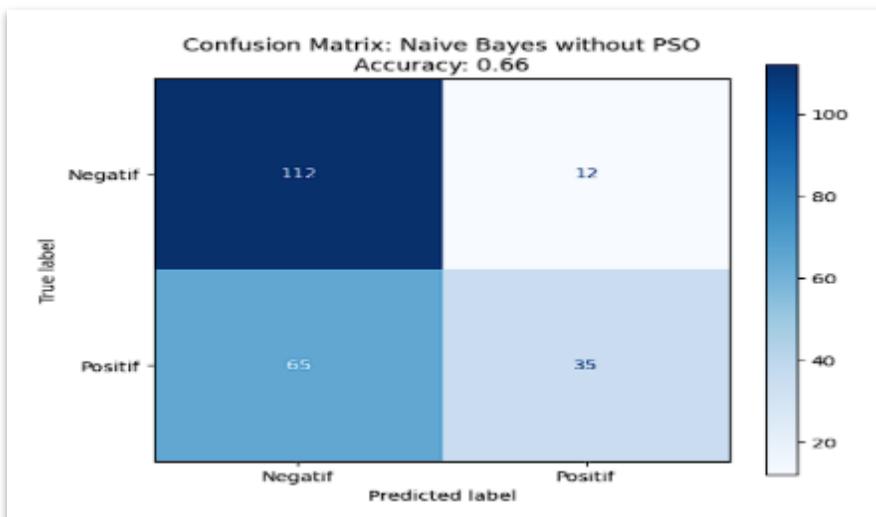
BAB V

HASIL DAN PEMBAHASAN

5.1 Evaluasi Model

5.1.1 Naïve Bayes Tanpa Particle Swarm Optimization (PSO)

Pada sub-bab ini akan ditampilkan hasil evaluasi terhadap model yang telah dibuat dan dilatih sebelumnya. Pada pengujian pertama ini dilakukan hanya menggunakan model dari algoritma Naïve Bayes tanpa menggunakan seleksi fitur, yang dimana hasil pada pengujian ini untuk mengetahui tingkat keakuratan model klasifikasi dan hasil dari pengujian ini nantinya akan dibandingkan dengan pengujian selanjutnya. Adapun data yang digunakan untuk menguji model adalah data sentimen “Ibu Kota Nusantara” pada twitter yang telah dibagi sebelumnya. Hasil performa model ditampilkan dalam bentuk heatmap. Adapun penilaian performa model dapat dilihat menggunakan confusion matrix. Pada confusion matrix dapat diketahui beberapa informasi yaitu akurasi prediksi dari hasil uji model. Pengujian pertama dilakukan pada model dengan presentase data latih 80% dan 20% data uji. Dengan jumlah *dataset* sebanyak 1116 *tweet* didapatkan hasil sebagai berikut



Gambar 5.1 : Visualisasi Confusion Matrix: Naive Bayes tanpa PSO

Pada gambar 5.1 dapat ditarik beberapa informasi yaitu :

- *True Negative* (TN) : Data berlabel negatif yang diprediksi benar sebagai label negatif sebanyak 112
- *True Positive* (TP) : Data berlabel positif yang diprediksi benar sebagai label positif sebanyak 35
- *False Negative* (FN) : Data berlabel positif yang diprediksi salah sebagai label negatif sebanyak 65
- *False Positive* (FP) : Data berlabel negatif yang diprediksi salah sebagai label positif sebanyak 12
- Hasil *Accuracy* Naïve Bayes tanpa PSO yang didapatkan sebesar 66%

Kemudian, peneliti melakukan perhitungan secara manualisasi untuk menghitung nilai performa dari model *naive bayes* pada nilai akurasi yang dihitung dari tabel *multiclass confusion matrix* di atas. Berikut proses perhitungan nilai akurasi secara manual

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

$$Akurasi = \frac{35 + 112}{35 + 112 + 12 + 65} \times 100$$

$$Akurasi = \frac{147}{224} = 0,66 \times 100$$

$$\boxed{Akurasi = 66\%}$$

5.1.2 Evaluasi Model Naïve Bayes Dengan Particle Swarm Optimization (PSO)

Pengujian berikutnya adalah pengujian dengan menggunakan model Naïve Bayes dan menambahkan seleksi fitur dari *Particle Swarm Optimization* (PSO). Hasil dari pengujian ini adalah untuk mengetahui tingkat keakuratan suatu model

klasifikasi yang berbasis seleksi fitur, yang dimana hasil dari pengujian tersebut akan dibandingkan dengan pengujian sebelumnya.

Pada pengujian kedua ini, peneliti menggunakan dataset yang digunakan pada pengujian pertama yang telah mendapatkan nilai akurasi 66%. Eksperimen yang dilakukan pada pengujian kedua ini menggunakan Parameter Population Size, maximum number of generation dan Inertia Weight. Penggunaan nilai parameter tersebut bertujuan untuk mencari manakah hasil akurasi yang terbaik.

1. Population Size

Dalam eksperimen ini, parameter *Population Size* digunakan untuk menentukan jumlah partikel dalam algoritma *Particle Swarm Optimization* (PSO). Setiap partikel dalam PSO mewakili solusi kandidat yang berbeda, dalam hal ini, kombinasi fitur yang berbeda untuk model Naive Bayes. Menggunakan berbagai ukuran populasi bertujuan untuk mengeksplorasi bagaimana jumlah partikel mempengaruhi kemampuan PSO dalam menemukan subset fitur yang optimal untuk meningkatkan akurasi model Naive Bayes.

Eksperimen awal pada pengujian kedua ini adalah meragamkan nilai pada Parameter *Population Size* dari 1-10 dengan nilai *Inertia Weight* 1.0, dan *Maximum Number of Generation* 50 bernilai tetap.

Hasil dari percobaan ini menunjukkan nilai *Accuracy* pada tabel dibawah ini

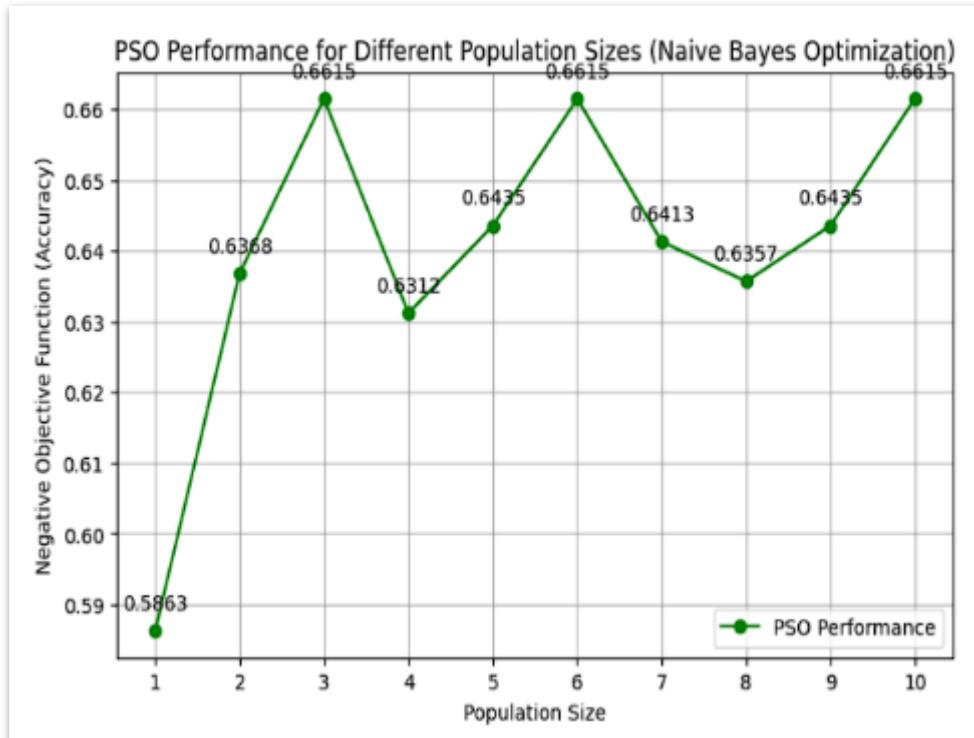
Tabel 5.1 : Hasil Eksperimen Algoritma Naïve Bayes berbasis PSO dengan nilai Parameter Population Size 1-10

<i>Population Size</i>	<i>Accuracy</i>
1	58,63%
2	63,68%
3	66,15%
4	63,12%
5	64,35%

6	66,15%
7	64,13%
8	63,57%
9	64,35%
10	66,15%

Hasil dari uji coba dengan meragamkan nilai pada Parameter Population Size dari 1-10 dengan nilai *inertia weight* dan *Maximum Number of Generation* yang tetap didapatkan hasil tertinggi pada nilai Parameter Population Size 3, 6, dan 10 dengan nilai Accuracy sebesar 66,15%.

Adapun visualisasi dari Population Size 1-10 pada grafik di bawah ini



Gambar 5.2 : Visualisasi Population Size 1-10

Dari hasil dari uji coba pertama dengan nilai Parameter Population tertinggi peneliti mengambil Population Size 10 karena dengan populasi yang lebih besar, lebih banyak partikel yang terlibat dalam proses pencarian solusi.

Ini meningkatkan kemampuan algoritma untuk menjelajahi ruang solusi secara lebih luas, sehingga peluang untuk menemukan solusi optimal menjadi lebih besar.

2. *Maximum Number of Generation*

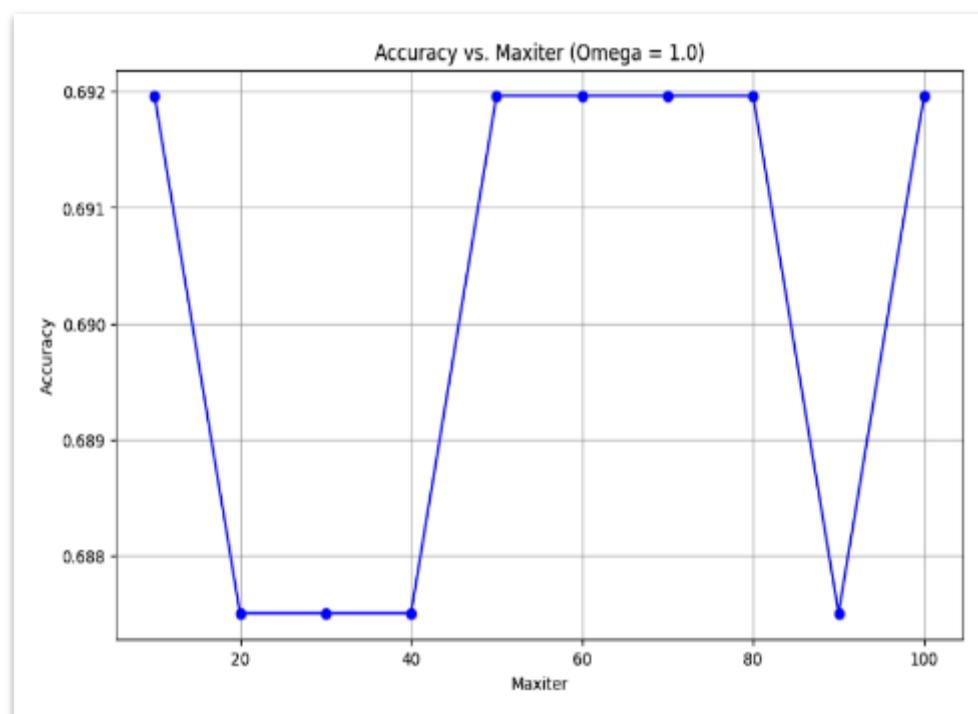
Selanjutnya parameter *Maximum Number of Generation* dalam algoritma *Particle Swarm Optimization* (PSO) dapat menentukan jumlah iterasi yang akan dijalankan oleh algoritma. Setiap iterasi memungkinkan partikel-partikel dalam PSO untuk memperbarui posisi mereka dalam ruang solusi, mengeksplorasi kombinasi fitur yang berbeda. Dengan mengubah nilai parameter ini, peneliti bertujuan untuk memahami bagaimana jumlah iterasi mempengaruhi kemampuan PSO dalam menemukan subset fitur yang optimal untuk meningkatkan akurasi model Naive Bayes. Pada percobaan kedua yang dimana pengujian dilakukan dengan meragamkan nilai parameter *Maximum Number of Generation* dari 10-100 atau berkelipatan 10 dengan nilai parameter pada *Inertia Weight* 1,0 yang tetap. Hasil dari percobaan kedua dapat dilihat pada tabel di bawah ini.

Tabel 5.2 : Hasil Eksperimen Algoritma Naïve Bayes berbasis PSO dengan nilai Parameter *Maximum Number of Generation* 10-100 atau berkelipatan 10

Maximum Number of Generation	Accuracy
10	69,19%
20	68,75%
30	68,75%
40	68,75%
50	69,19%
60	69,19%
70	69,19%
80	69,19%
90	68,75%
100	69,19%

Hasil dari uji coba kedua dengan meragamkan nilai pada parameter *Maximum Number of Generation* 10-100 atau berkelipatan 10 dengan Population Size 10 dan Inertia Weight 1.0 didapatkan hasil tertinggi pada nilai parameter *Maximum Number of Generation* 10, 50, 60, 70, 80, dan 100 dengan hasil nilai Accuracy sebesar 69,19%.

Adapun visualisasi dari Maximum Number of Generation ditampilkan pada grafik dibawah ini.



Gambar 5.3 : Visualisasi Maximum Number Of Generation

Dari hasil uji coba kedua dengan nilai Parameter *Maximum Number of Generation* tertinggi peneliti mengambil jumlah maksimum generasi yang lebih tinggi, yaitu 100 karena memberikan algoritma lebih banyak kesempatan untuk mengeksplorasi ruang solusi secara mendalam. Ini meningkatkan peluang untuk menemukan solusi yang lebih baik atau memastikan bahwa solusi yang ditemukan benar-benar optimal.

3. *Inertia Weight*

Pada eksperimen ketiga dengan parameter *Inertia Weight* dapat menentukan seberapa besar pengaruh kecepatan sebelumnya terhadap kecepatan baru dari partikel. Nilai *Inertia Weight* yang optimal dapat membantu partikel-partikel dalam PSO untuk mengeksplorasi ruang solusi secara efektif tanpa terlalu cepat terkonsentrasi pada solusi lokal. Dengan mengatur nilai *Inertia Weight* yang sesuai, PSO dapat membantu meningkatkan akurasi model Naive Bayes dengan menemukan kombinasi fitur yang lebih relevan. Eksperimen ini bertujuan untuk memahami bagaimana variasi nilai Inertia Weight mempengaruhi kemampuan PSO dalam meningkatkan akurasi Naive Bayes.

Pada percobaan ketiga dengan meragamkan nilai pada parameter Inertia Weight dari 1.0 - 2.0 dengan nilai Parameter Population Size 10 dan Maximum Number of Generation 100. Hasil dari percobaan ketiga dapat dilihat pada tabel di bawah ini.

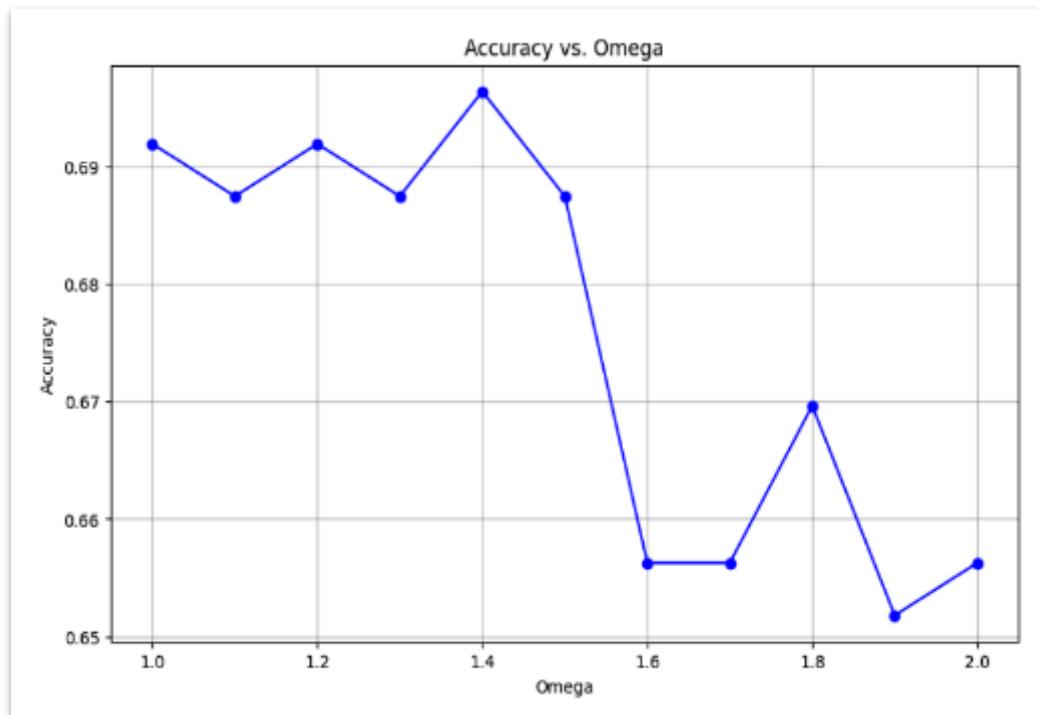
Tabel 5.3 : Hasil Eksperimen Algoritma Naïve Bayes berbasis PSO dengan nilai parameter Inertia Weight 1.0 - 2.0

<i>Inertia Weight (omega)</i>	<i>Accuracy</i>
1.0	69,19%
1.1	68,75%
1.2	69,19%
1.3	68,75%
1.4	69,64%
1.5	68,75%
1.6	65,62%
1.7	65,62%
1.8	66,96%
1.9	65,17%

<i>Inertia Weight (omega)</i>	<i>Accuracy</i>
2.0	65,62%

Hasil dari uji coba ketiga dengan meragamkan nilai parameter Inertia Weight didapatkan hasil akurasi tertinggi terdapat di Inertia Weight 1.4 dengan nilai Accuracy 69,64%.

Adapun visualisasi dari *Inertia Weight* ditampilkan pada grafik dibawah ini.



Gambar 5.4 : Visualisasi Inertia Weight (omega)

Dari hasil percobaan pertama, percobaan kedua, dan percobaan ketiga pada eksperimen Naïve Bayes berbasis fitur Particle Swarm Optimization (PSO) dengan merubah nilai parameter pada Population Size 1-10, Maximum Number of Generation 10-100, dan nilai Inertia Weight 1.0 – 2.0, didapatkan hasil akurasi tertinggi pada saat nilai Parameter Population Size 10, Maximum Number of Generation 100 dan Inertia Weight 1.4 dengan nilai Accuracy sebesar 69,64%.

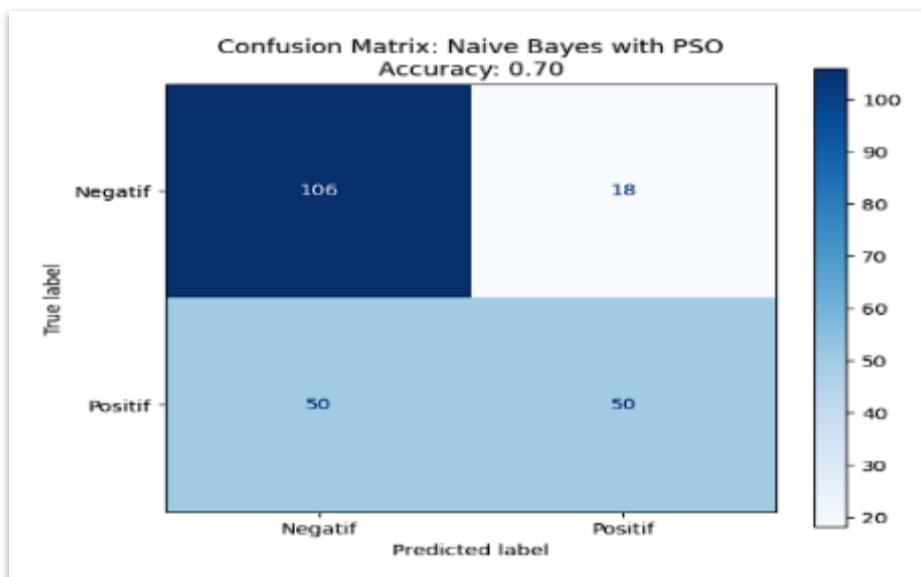
Berikut tabel hasil dari eksperimen Pengujian Naive Bayes dengan PSO menggunakan paramater *Population Size*, *maximum number of generation* dan *Inertia Weight*.

Tabel 5.4 : Hasil Eksperimen pada Algoritma Naïve Bayes berbasis PSO

No	Parameter	Nilai
1	<i>Population Size</i>	10
2	<i>Maximum Number Of Generation</i>	100
3	<i>Inertia Weight</i>	1.4

Dari tabel 5.4 dapat disimpulkan bahwa eksperimen Naïve Bayes berbasis fitur *Particle Swarm Optimization* (PSO) dengan nilai parameter pada *Population Size* 10, *Maximum Number of Generation* 100, dan nilai *Inertia Weight* 1.4 dapat meningkatkan nilai Accuracy sebesar 69,64%.

Untuk melihat lebih lanjut hasil dari Confusion Matrix pada pengujian Naïve Bayes berbasis fitur Particle Swarm Optimization (PSO) dapat dilihat pada heatmap berikut.



Gambar 5.5 : Hasil Confusion Matrix: Naive Bayes dengan PSO

Pada gambar 5.5 dapat ditarik beberapa informasi yaitu :

- *True Negative* (TN) : Data berlabel negatif yang diprediksi benar sebagai label negatif sebanyak 106
- *True Positive* (TP) : Data berlabel positif yang diprediksi benar sebagai label positif sebanyak 50
- *False Negative* (FN) : Data berlabel positif yang diprediksi salah sebagai label negatif sebanyak 50
- *False Positive* (FP) : Data berlabel negatif yang diprediksi salah sebagai label positif sebanyak 18
- Hasil *Accuracy* Naïve Bayes dengan PSO yang didapatkan sebesar 70%

Selanjutnya peneliti melakukan perhitungan secara manual untuk menghitung nilai performa dari model naive bayes dengan penerapan fitur *Particle Swarm Optimization* PSO pada nilai akurasi yang dihitung dari tabel confusion matrix di atas. Berikut proses perhitungan nilai akurasi secara manual

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \times 100$$

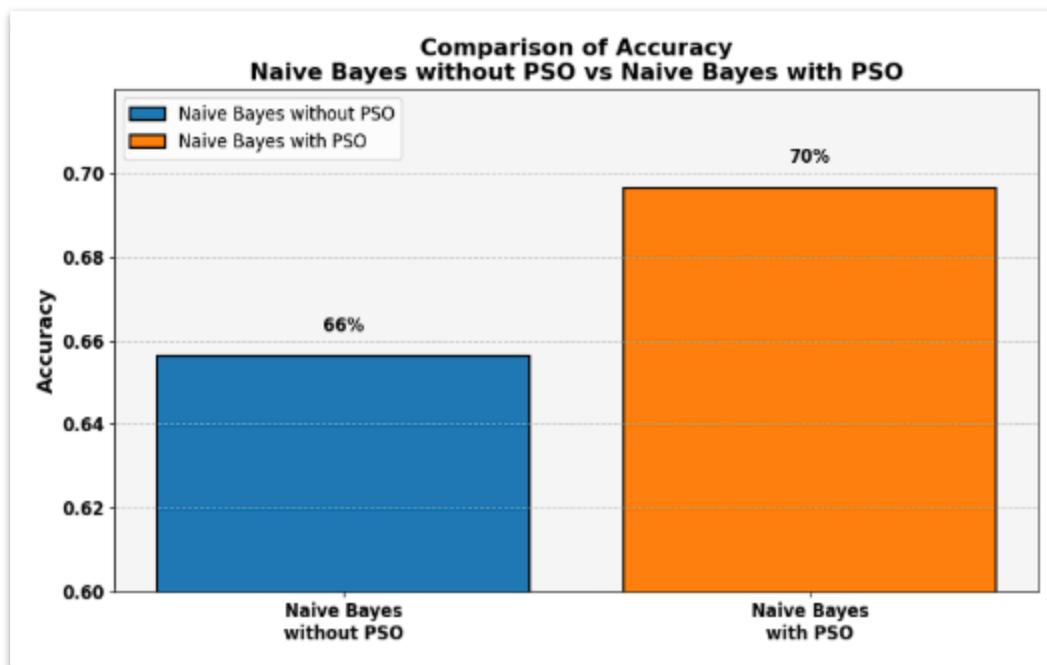
$$Akurasi = \frac{50 + 106}{50 + 106 + 18 + 50} \times 100$$

$$Akurasi = \frac{156}{224} = 0,70 \times 100$$

$$Akurasi = 70\%$$

5.2 Analisis Hasil Akurasi

Seberapa dekat nilai prediksi dari nilai sebenarnya merupakan bukti bahwa performa dari sebuah algoritma memiliki hasil yang akurat dalam mengklasifikasi sentimen. Dari pengujian yang telah dilakukan dengan menggunakan Naïve Bayes beserta optimasi menggunakan seleksi fitur Particle Swarm Optimization menghasilkan performa atau nilai akurasi terbaiknya. Berikut adalah grafik dari perbandingan hasil akurasi pengujian 1 dan pengujian 2 yang telah dilakukan.



Gambar 5.6 : Grafik dari Perbandingan Nilai Accuracy

Berdasarkan dari gambar grafik pada gambar 5.3, telah menunjukkan bahwa terdapat peningkatan dari hasil Accuracy antara pengujian 1 dan pengujian 2. Terdapat peningkatan nilai Accuracy sebesar 4% pada pengujian kedua yang awalnya hanya 66% menjadi 70%, peningkatan nilai Accuracy tersebut dikarenakan pada pengujian kedua menggunakan Naive Bayes berbasis fitur *Particle Swarm Optimization*. Hal ini membuktikan bahwa seleksi fitur Particle Swarm Optimization (PSO) kurang maksimal dalam meningkatkan dan mengoptimalkan nilai akurasi Naive Bayes dalam melakukan klasifikasi sentimen.

Dari hasil eksperimen yang dilakukan, dapat disimpulkan bahwa optimasi hyperparameter menggunakan algoritma PSO kurang maksimal dalam

meningkatkan kinerja model Naive Bayes. Hal ini terlihat dari peningkatan nilai akurasi setelah optimasi. Meskipun peningkatan ini tidak terlalu signifikan, namun menunjukkan bahwa PSO efektif dalam mencari nilai optimal untuk hyperparameter alpha.

Penelitian ini memberikan bukti bahwa metode optimasi seperti PSO dapat digunakan untuk meningkatkan performa model machine learning meskipun kurang maksimal. Selain itu, hasil ini menunjukkan bahwa dengan penggunaan PSO, model Naive Bayes dapat menjadi lebih andal dalam klasifikasi sentimen.

BAB VI

PENUTUP

6.1 Kesimpulan

Pada penelitian ini penulis telah melakukan pemodelan menggunakan metode Naïve Bayes classifier dan seleksi fitur Particle Swarm Optimization (PSO), maka berdasarkan hasil dari penelitian yang telah dilakukan dapat disimpulkan bahwa analisis sentimen dapat diterapkan menggunakan metode Naïve Bayes serta penambahan seleksi fitur *Particle Swarm Optimization* (PSO).

Pada penelitian ini dengan menggunakan dataset 1116 *tweet* dan komposisi data 80% : 20%. Pengujian ke-1 menggunakan algoritma Naïve Bayes mendapatkan nilai *Accuracy* sebesar 66%,. Pengujian ke-2 dengan menambahkan seleksi fitur menggunakan *Particle Swarm Optimization* (PSO) dengan merubah nilai parameter pada Population Size 1-10, Maximum Number of Generation 10-100, dan nilai Inertia Weight 1.0 – 2.0, didapatkan hasil akurasi tertinggi pada saat nilai Parameter Population Size 10, Maximum Number of Generation 100 dan Inertia Weight 1.4 dengan nilai *Accuracy* sebesar 70%. Peningkatan nilai akurasi yang didapatkan pada pengujian ke-2 sebesar 4%, dengan membuktikan bahwa penerapan seleksi fitur PSO kurang maksimal dalam meningkatkan dan mengoptimalkan nilai akurasi Naïve Bayes dalam mengklasifikasi sentimen.

6.2 Saran

Penelitian ini masih memiliki banyak kekurangan, oleh sebab itu penulis memberikan saran untuk penelitian selanjutnya, yaitu:

1. Mengkombinasikan atau membandingkan PSO dengan teknik optimasi lain seperti *Genetic Algorithms* (GA), *Simulated Annealing* (SA), dan *Differential Evolution* (DE) untuk mengetahui metode mana yang lebih efektif dalam konteks optimasi parameter Naive Bayes.
2. Menggunakan lebih banyak jenis data, seperti forum diskusi atau review di platform media sosial lainnya, selain hanya menggunakan data dari Twitter.

3. Mempertimbangkan metode pelabelan *dataset* secara manual dan dengan begitu dapat membandingkan hasil keakuratan dari pelabelan *dataset* secara manual dan otomatis.
4. Menggunakan metode analisis sentimen lain seperti *Support Vector Machine* (SVM), K-NN, Decision Tree dan dengan begitu dapat membandingkan hasil performa yang dilakukan untuk mencari metode klasifikasi yang terbaik.

DAFTAR PUSTAKA

- [1] S. Lestari *dkk.*, “Analisis Sentimen Masyarakat Indonesia terhadap Pemindahan Ibu Kota Negara Indonesia pada Twitter,” vol. 8, no. 1, hlm. 13–22.
- [2] Syahril Dwi Prasetyo, Shofa Shofiah Hilabi, dan Fitri Nurapriani, “Analisis Sentimen Relokasi Ibukota Nusantara Menggunakan Algoritma Naïve Bayes dan KNN,” *Jurnal KomtekInfo*, hlm. 1–7, Jan 2023, doi: 10.35134/komtekinfo.v10i1.330.
- [3] A. Kurniawan dan S. Waluyo, “Penerapan Algoritma Naive Bayes Dalam Analisis Sentimen Pemindahan Ibukota Pada Twitter Application Of Naive Bayes Algorithm In Capital Movement Sentiment Analysis On Twitter,” 2022. [Daring]. Tersedia pada: <https://senafti.budiluhur.ac.id/index.php>
- [4] S. Oleh, “ANALISIS SENTIMEN MASYARAKAT DI TWITTER TERHADAP AKSESIBILITAS DISABILITAS DI RUANG PUBLIK MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM) DENGAN KERNEL RADIAL BASIS FUNCTION (RBF) Mitha Rachma Putri.”
- [5] E. Indrayuni dan A. Nurhadi, “OPTIMASI NAIVE BAYES BERBASIS PSO UNTUK ANALISA SENTIMEN PERKEMBANGAN ARTIFICIAL INTELLIGENCE DI TWITTER,” *INTI Nusa Mandiri*, vol. 18, no. 1, hlm. 65–70, Agu 2023, doi: 10.33480/inti.v18i1.4282.
- [6] C. F. Hasri dan D. Alita, “PENERAPAN METODE NAÏVE BAYES CLASSIFIER DAN SUPPORT VECTOR MACHINE PADA ANALISIS SENTIMEN TERHADAP DAMPAK VIRUS CORONA DI TWITTER,” *Jurnal Informatika dan Rekayasa Perangkat Lunak (JATIKA)*, vol. 3, no. 2, hlm. 145–160, 2022, [Daring]. Tersedia pada: <http://jim.teknokrat.ac.id/index.php/informatika>
- [7] H. B. Jatmiko, N. T. Kurniadi, dan D. Maulana, “Optimasi Naïve Bayes Dengan Particle Swarm Optimization Untuk Analisis Sentimen Formula E-Jakarta Optimization of Naïve Bayes With Particle Swarm Optimization for Sentimen Analysis of Jakarta E-Prix,” 2022.
- [8] R. Darmawan dan A. Surahmat, “Optimalisasi Support Vector Machine (SVM) Berbasis Particle Swarm Optimization (PSO) Pada Analisis Sentimen Terhadap Official Account Ruang Guru Di Twitter,” 2022. [Daring]. Tersedia pada: <http://ejurnal.ubharajaya.ac.id/index.php/JKI>

- [9] N. Legiawati, T. I. Hermanto, dan Y. R. Ramadhan, “Analisis Sentimen Opini Pengguna Twitter Terhadap Perusahaan Jasa Ekspedisi Menggunakan Algoritma Naïve Bayes Berbasis PSO,” *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 4, hlm. 930, Agu 2022, doi: 10.30865/jurikom.v9i4.4629.
- [10] S. Dwi, “SENTIMENT ANALYSIS FOR TWITTER USING MULTINOMINAL NAIVE BAYES A THESIS Presented as Partial Fullfillment of Requirements to Obtain Sarjana Komputer Degree in Informatics Engineering Department.”
- [11] I. Kurniawan, “SKRIPSI ANALISIS SENTIMEN VAKSIN COVID-19 DI INDONESIA.”
- [12] Gading Teguh Santoso, “ANALISIS SENTIMEN PADA TWEET DENGAN TAGAR #BPJSRASARENTENIR MENGGUNAKAN METODE SUPPORT VECTORE MACHINE (SVM).”
- [13] Irma Surya Kumala Idris, Yasin Aril Mustofa, dan Irvan Abraham Salihi, “Analisis Sentimen Terhadap Penggunaan Aplikasi Shopee Menggunakan Algoritma Support Vector Machine (SVM),” *J Inf Sci*, vol. 36, no. 6, hlm. 823–848, Des 2010, doi: 10.1177/0165551510388123.
- [14] C. Mohi, “ANALISIS SENTIMEN PADA TWEETS DIVISI HUMAS POLRI DENGAN METODE NAÏVE BAYES CLASSIFIER.”
- [15] “SKRIPSI ANALISIS SENTIMEN MASYARAKAT PADA MEDIA.”
- [16] L. Yusuf dan S. Masripah, “SENTIMEN ANALISIS CHATGPT DENGAN ALGORITMA NAÏVE BAYES DAN OPTIMASI PSO,” *INTI Nusa Mandiri*, vol. 18, no. 1, hlm. 59–64, Agu 2023, doi: 10.33480/inti.v18i1.4230.

LAMPIRAN

No.	Text
1	ini fix sih orang miskin gak bakal bisa masuk apalagi yg niat jualan di emperan IKN sekalipun pendukung mati an si wowo
2	IKN bakal mangkrak mudah han cepet mangkrak
3	Sama kaya pembuatan IKN pejabatnya ingin buat lingkungan hunian mewah duitnya dari pajak masyarakat Sesederhana itu
4	IKN emg pemerintah pd mau pindah Apalagi pejabat nya
5	Apakah IKN di bangun untuk menjadi kota tertutup dan hanya orang tertentu saja yg boleh masuk
6	oh opiniku semakin membaik soal ikn udah ada kang gorengan gaes
7	wakakkak trbiasa rame dn seru bng di jkt klo di kalimtan msh bnyak hutan lebat klo spengthuan sya IKN memng bng luas sekali dn mngkn prlu thnn lgi bru bsa cukp rame
8	Ada dugaan bhw Pantai Indah Kapuk PIK tsb ke depannya menjadi Singapura di Indonesia yg Terkait IKN ke Kalimantan Waspada dan Stoplah aset gedung kantor dll yg di DKI Jakarta di jual pd pihak ke atau pd oligarki Aseng Asing aroma komunis Mencegah lbh baik drpd di tindas
9	IKN gorengan biji
10	Khusus watimpres nanti harus tinggal di IKN
11	pernah ada denger kejadian serupa jd di sini sumpah semenjak kerja di luar jawa makin makin melek sm ketimpangan infrastruktur akses jalan banyak yg ga layak yaAllah nangis pemerentahnya malah megahin IKN anjrit dosa bgt
12	Dia lupa kalo yang bakal mayoritas di IKN itu adalah PNS Umbi umbian ya rakyat juga itungannya
13	Anak Abah bahas IKN
14	Very easy Krn kekuatan politik jokowi di parlemen tak ada Jika publik mocking soal makan siang gratis yang entah jadi atau nggak itu mau tak mau IKN yg harus dikorbankan dan konsekwensi lanjutannya mangkrak Jika mangkrak jokowi ke KPK
15	Contohnya warlok disekitar IKN
16	Penduduk ikn no comment ah yang enjoy jakarta aja yang boleh komen
17	Bukankah IKN juga produk Makin aneh aja bernegara ini
18	Jalan tengah malam ke kota sebelah biar sampai pas di pagi hari eh selama perjalanan malah barengan sama truk IKN yang besar semua anjir Ô¶è
19	sumpah gweh sngt syg sm mu awas klo sakitin hati moengil ku ini lo akan membangkitkan raja iblis
20	Pajak¬s dinaikin diada in BBM lustrik pam benaran ajiib sembako lauk pauk meroket jajan di resto juga makin mehong Jogetin gemoy aja daahhhh Keringet rakyat diperezzz buat mbayarin hutang jatuh tempo dan keberlanjutan ikn
21	smoga bukan di ikn ya kak
22	Ini probably akan terjadi btw Arah nya sudah ke sana Sampai ada manuver narasi kapan hari yang menyatakan maksi gratis lebih murah vs IKN Mengorbankan jokowi adalah salah satu cara paling realistik untuk menstabilkan kondisi politik at least in the first years
23	Cocok suruh aja Gibran Nyepi jaga IKN bersama bapaknya dan kroninya disana biar jarang dilihat orang Kalau di Jkt ntar jd sasaran demo JKT dan istana negara biar diurus Prabowo jd tdk risi ditunggui si Samsul

No.	Text
24	Realistically speaking akan better bagi prabsky untuk membiarkan IKN mangkrak Krn bebannya nanti nggak cuma anggaran tapi jd transisi birokrasi Berat sekali untuk konsolidasi pemerintahan baru Lebih murah dan mudah fokus ke makan siang gratis yang memang janjinya
25	Bandara hanya utk VVIP bkn utk rakyat jelas IKN untuk siapa
26	dah dibilang ikn bkn utk bersama wkwkw
27	Negara rugi nangkep trus ngasih makan slama di prodeo Mending kirim kerja Rodi Romusha bangun rel di Sumatera Kalimantan Sulawesi atau bikin jln di Papua atau jd kuli buat bangun IKN
28	Benar pak Pelayanan kesehatan masyarakat yg masih kurang diseluruh pelosok tanah air ini sungguh jauh lebih penting daripada proyek IKN atau kereta Whoosos
29	Just clocked rumah dinas baru menteri di IKN deket hutan bekas hutan
30	Amanah undang orang miskin dan anak terlantar d pelihara negara ini malah bangun ikn
31	Ya ya ya gambar yang lumayan Tapi itu tidak cukup untuk memungkiri fakta bahwa habitat satwa endemik terus terancam rusak dan seutuhnya hilang dirampas pembangunan IKN salah satunya Saya bantu ingatkan apa kata menteri kalian
32	Anggota Dewanpun baru mau tinggal di IKN klo infrastruktur prasarana dn smua fasilitas udah lngkap semua Pdhal nunggu lngkap itu baru bisa di th Alasan aj padune ora gelem pindah
33	Ikn apa buat program makan siang gretong atau buat rumah menteri yg di ikn yg masih di komenin krn blm kliatan wah
34	Liat ini para buzzer ikn tll
35	Kalo emang beneran gunung di Malaysia gaada yg diledakin Gaada tanah longsor krn penebangan liar Gaada pejabat korup krn yg korup pemegang proyek IKN tersesat di alam Jin UPS Jin emang ada tapi tdk se hiperbola itu Orang dulu kan gabut ja
36	lho katanya rakyat ga makan infrastruktur giliran ada kasus gini nyalahin IKN
37	Keluarga saya sangat suka mkn ikan walaupun kandang hanya bisa beli ikan pindang tapi hampir tiap hari diusahakan bisa mkn ikn Yahh hmm tapi kadang ngga ke beli juga ,Ã¢Â¶,Ã¢Â¶
38	aja tugas di IKN daripada ga berguna di ISTANA
39	Biar IKN yang nempati para oposisi
40	Menurut gw kurang kurangin deh berita kayak gini secara g langsung kalian nyakinin perasaan masyarakat indo yg memperjuangkan kemajuan pembangunan IKN utk elit negara dibandingkan pembangunan infrastuktur mikro untuk masyarakat ekonomi menengah ke bawah
41	urusan IKN apa yang sudah menyusahkanmu arka
42	Jangan berharap pd Rezim Ngurus makan siang gratis saja sulit Sekarang Mukijo sdg gundah mikir IKN mikir cawe anaknya mau ikut Pilkada mikir cawe Makan Siang gratis pusing dia Kalau sdg stress perginya ke Pasar mana berani ke
43	Uu ikn kalau tdk dilanjutkan cabut dulu uu ikn di dpr atau perpu ikn rasanya tdk mungkin Ttp biaya ikn bisa di kurangi dulu di apbn utk pos makan gratis Ikn bukan hanya pakai apbn ttp juga dari suasta
44	Pembangunan IKN Dan Visi Masa Depan Bangsa Dalam Model Dan Paradigma Komunikasi Pembangunan Geby PO Bus OBnya SMK Lingga Kencana Depok Pendek Bobon
45	Mentan Sebut Sulawesi Barat Bakal Jadi Pemasok Bahan Pangan IKN

No.	Text
46	Pesona Gua Tapak Raja Menjadi Unggulan Daya Tarik Wisata Di IKN SMK Lingga Kencana Depok PO Bus Geby OBnya Bobon Dilan
47	Presiden Berkantor Di IKN Juli Smart Defense System Diterapkan PO Bus Casemiro OBnya Geby Pendek Garnacho
48	Trotoar Untuk Pejalan Kaki Di IKN Dilengkapi Jalur Sepeda Tanaman Hijau Dan Ramah Disabilitas PO Bus Geby Dilan Senin SMK Lingga Kencana Depok OBnya
49	Menembus Terowongan Multi Utilitas IKN Rapi Dan Dilapisi Epoksi PO Bus Geby OBnya SMK Lingga Kencana Depok Love Wins Pendek
50	Pemerataan pembangunan itu tugas pemerintah pusat Solusinya apa Yaj betul bikin IKN Wkwkwk
51	Gedung dpr segera selesai di ikn yg lama di jkt dijual utk jadi mal
52	Ikn jadi kota modern Terus kerjakan sampai selesai sesuai amanat uu ikn
53	IKN dah berada pulau nusa Kambangan saja tapi yang ini khusus kerah putih
54	Rame ogah pindah ke IKN mau tetap di Jakarta Prabowo juga ogah an pindah IKN mau tetap di Jakarta Jadi apa jadinya nanti
55	Kalau IKN mangkrak bakal banyak yg terjerat hukum karena banyak menghabiskan duit negara
56	Rezim mementingkan IKN titik
57	Sapa yg sudi tinggal di IKN
58	Orang orangan pinterr seperti dan serta yg sejenisnya yg cucokkk pindah ke IKN baca ai ke en
59	IKN itu bagian proyek cina jalur Suster silk road belt
60	Mukidov El Bansos saja pensiun pulangnya ke Solo karena kalau ke IKN tidak bisa terus cawe cawe demi dinasti konyol politik
61	Sabar bu uang negara sedang fokus membangun IKN demi kemajuan bangsa dan negara ini
62	RK mah sdh pny jbtn kurator IKN sdh sibuk jakarta berik
63	YaAllah kalo emg jalanku disini permudahlah tp kalo dipindah ke IKN yaudahlah
64	Pembangunan IKN menurut Prabowo Subianto akan tetap melindungi serta menjamin kesejahteraan masyarakat lokal dan masyarakat adat yang berada di IKN
65	Lindungi Masyarakat Lokal Menanggapi dugaan kurangnya perhatian terhadap masyarakat lokal dalam pembangunan IKN Prabowo Subianto menegaskan komitmennya untuk melindungi kepentingan warga lokal termasuk masyarakat adat
66	Kepada media asal Timur Tengah ini Prabowo Subianto menegaskan komitmennya untuk melanjutkan proyek IKN yang sudah dimulai di era pemerintahan Presiden Joko Widodo ini
67	Bukan hanya untuk pengembangan wilayah Kalimantan Timur tempat IKN berada pemindahan ibu kota juga penting untuk keberlangsungan Jakarta
68	Presiden terpilih Prabowo Subianto kembali menegaskan sikap dan komitmennya untuk mendukung dan melanjutkan pemindahan ibu kota ke Ibu Kota Nusantara IKN
69	Wawancara Dengan Media Timur Tengah Prabowo Subianto Tegaskan Komitmen Untuk Lanjutkan Pembangunan IKN
70	Bapak harus cermati mana yang lebih prioritas sekarang ini meningkatkan kesejahteraan rakyat atau membangun IKN Pikirkan disitu
71	Untuk satu keluarga Mukidi aja Biar enak tuh luas kan buat mereka guling an jgn lupa pemuja dan para fans nya suruh pindah aja ke IKN
72	Indonesia itu banyak beruntung blessing in disguise Proyek ngawur dengan kebocoran bisa dikemudian hari berguna Termasuk IKN meskipun dikerjakan

No.	Text
	dg ngawur tapi itu memang satu satunya jalan untuk mengatasi kepadatan Jawa yg sdh di luar ak
73	Dalam wawancara tersebut Prabowo juga menegaskan komitmennya untuk melindungi serta menjamin kesejahteraan masyarakat adat yang berada di IKN
74	Presiden RI terpilih periode Prabowo Subianto kembali menegaskan dukungan serta komitmennya untuk melanjutkan kebijakan pembangunan Ibu Kota Nusantara IKN yang digagas oleh Presiden Joko Widodo Jokowi
75	Wawancara dengan Al Jazeera Prabowo Tegaskan Dukungan untuk Kelanjutan Pembangunan IKN
76	Pembangunan Ibu Kota Negara IKN Nusantara harus membuat warga Kaltim tambah sejahtera
77	Tol IKN di Kaltim Siap Fungsional Sebelum HUT RI Agustus Progres Pembangunan Capai Persen lewat
78	Mungkin mau pindah Bu tapi ya itu naikin salary dll nya Nanti kan ada bandara khusus IKN JKT dejat lah
79	Otorita Ibu Kota Nusantara IKN dan Kemenkes bersatu dalam pelatihan dan pembinaan keamanan pangan siap saji bagi pelaku UMKM Tong tji geby bu bos panjang anggi obnya hujan di bali subang aurora Senin siang udah
80	Yang suka bawa mandau paling banter jadi satpam bandara Mungkin kata kata gua ini menyakitkan tapi beneran kok lihat saja lima tahun kedepan penduduk asli sekitar IKN ada dimana Coba tebak dimana sekarang betawi pemilik tanah tanah sekitar SCD
81	Kedok aja ikn aslinya cari cuan dari pembangunannya kadung mlebu weteng wareg dan terus menerus sambung menyambung menjadi satu Weesss uangel kandanane dok
82	A karena A Lunga B Tuku C Dolan D Tuku B belakangnya goreng
83	Eitss mending bangun IKN dong
84	IKN cuma ladang buat bisnis korupsi doang
85	Nah kan moso akeh sing wira wiri ikn ga dimurahne
86	Faskes RS di pelosok dg fasilitas maju dan canggih harusnya lebih jd prioritas drpd IKN sygnya yg kyk gt gk bs diiklanin ke singapore
87	Gua kok nggak pingin amat tour ke IKN Krn pasti tdk ada Pasar grosir Tanah Abang tdk ada Sarinah
88	Lhoooo emang ada apa di IKN kan enak dpt fasilitas mewah Wewe gombel ditiup aja kalau parang melayang yaa lari lahh
89	Pengen bgt kek gini pola hidupnya sehat bgaaaa kamu Tapi apalah daya sekarang ak dah jadi anak kos jam tidur berantakan apalagi kl jadwal kuliah siang sampe sore Malem baru pulang lanjut ngerjain tugas sampe jam pagi Makanan pun ayam ge
90	Wrexham dibeli Ryan Reynolds &
91	Sudahlah biarkan saja Aslinya mereka juga bingung sendiri mau sampai mana kerjakan IKN Bisa jadi kerja tahunan bakal hancur itungan detik Kondisi lingkungan dan lahan sangat tidak mampu menopang sebuah kota besar
92	Orang yang membandingkan menyamakan Semangi yang visioner dengan yang proyek IKN adalah sebuah ketololan yang hakiki
93	IKN dari tempat jin buang anak sekarang diubah jadi tempat siluman aseng
94	Ngene Ki terus piye kok pada gak mau pindah trus utk siapa IKN
95	Pemerintah terus berupaya membangun sistem pertahanan cerdas atau smart defense system di IKN Nusantara Sistem pertahanan cerdas harus dibangun di IKN karena daerah tersebut nantinya akan menjadi pusat pemerintahan Indonesia

No.	Text
96	Mubadzir banget ikn kalo gaa dotempatin
97	Kenapa ogah pindah ke IKN bukankah proyek pemindahan IKN berdasarkan UU yang dibuat oleh Pemerintah &
98	Uangnya buat IKN ga sih buat apa menyenangkan rakyat prioritas negara itu pejabat
99	Semenjak kang ngibul ketahuan bohongnya gw jadi pesimis dg IKN itu proyek akal annya dia saja ternyata
100	Padahal ancaman ke depan bakal ada krisis pangan tp pemerintah sibuk sma IKN
101	memang pindah ke ikn bisa hidup abadi jakarta mau tenggelam kapan bro apa kamu ngak yg lebih dulu tenggelam ke dlm tanah di ikn banyak nyamuk malaria mending di jakarta banyak pencinta dunia malam yg bisa donor dar
102	Tunggu IKN
103	Ikn sih Napa jadi ikan
104	Warga negara kalimantan sadarlah IKN hanya kedok
105	wong LIBERALISASI dimulai dari sini kok SEMUA ANGGARAN KE IKN UNIVERSITAS URUS MASING MASING
106	Apa harus dipenajam kah Pdhl ada tanah negara yg tdk harus pakai ganti rugi tuk IKN nya
107	Pemerintah terus berupaya membangun sistem pertahanan cerdas di IKN Nusantara Sistem pertahanan cerdas harus dibangun di IKN karena daerah tersebut nantinya akan menjadi pusat pemerintahan Indonesia serta sesuai dengan rencana induk sistem pertahanan negara di IKN
108	Harusnya pindah ke IKN sepi ya pasti Tapi ada kenikmatan dalam kesepian bisa kongkalingkong dalam dalam banyak hal
109	th kedepan bisnis Travel &
110	Yg penting IKN jadi dulu Bu
111	Kalau presiden terpilih ga mau tinggal di IKN buat anak lanang aja
112	Buat san di IKN doang telan dana triliunan buat para panitia dan pejabat menginap di hotel berbintang miris uang negara dihambur-<kan buat yg ga penting
113	goblok anjeng ngentot itu duit IKN udah abis kah
114	Mulai detik ini gw tidak akan support IKN IKN ANJING
115	berapa puluh juta pedagang kecil yg bisa dibantu kalau dana ikn dialihkan buat mrk pak
116	kadang juga bingung kenapa kalo emang mau majuin indo harus bangun IKN Duit buat bangun tuh IKN padahal bisa buat yg lain ya meskipun ga segampang itu tapi kek buat apa anjir tuh IKN mana rumah dinas menterinya mahal bgt nilainya taik emang
117	Nunggu ikn beres ya gaiess
118	Menjelajahi Terowongan Multi Utilitas IKN Saluran yang Musnahkan Kabel Menggantung di Ibu Kota N
119	Mending semua pabrik pabrik di jabodetabek pindahkan ke ikn Biar berkurang jauh bebanya
120	Wah gak bisa booking lagi kalau di IKN kata anggota DPR Booking tiket maksudnya
121	AHY mau pindah ke IKN Krn sdh menikmati suasana di sana Ô¶èÔ¶èÔ¶è
122	Tinggal opsi Makan gratis berjalan tp konsekuensinya IKN mangkrak
123	IKN diKaltim bukan diKalteng Kalo IKN diKalteng maka saya &
124	Mari kita dukung pembangunan IKN
125	Wwkwkwk lahan di IKn jadi milik petani cina cek n ricek sendiri kesana

No.	Text
126	Andai duit IKN dipakai untuk perbaikan jalan berapa ribu KM jalan layak yg bisa dibuat
127	Pembangunan IKN Dan Visi Masa Depan Bangsa Dalam Model Dan Paradigma Komunikasi Pembangunan
128	Prabowo Tegaskan Komitmen untuk Kesejahteraan Masyarakat Adat di IKN
129	Pesona Gua Tapak Raja Menjadi Unggulan Daya Tarik Wisata Di IKN di Bali PO Bus OBnya Aurora Geby Garnacho
130	fix pindah ikn
131	Upaya Mengatasi Kesenjangan Proyek IKN Berhasil Tarik Minat Investasi Denmark di Bali PO Bus Geby SMK Lingga Kencana Depok Siang OBnya
132	Zero stunting bukan sekadar mimpi di IKN tapi komitmen nyata untuk menciptakan generasi penerus yang sehat dan tangguh
133	Klo ni org jd gubernur mak gua akan mengajukan pindah IKN
134	Jadi tukang parkir di IKN sepertinya cocok buat tukang parkir liar itu pan lokasinya masih dekat dengan alam liar
135	Trotoar Untuk Pejalan Kaki Di IKN Dilengkapi Jalur Sepeda Tanaman Hijau Dan Ramah Disabilitas Geby di Bali OBnya Heru Subang
136	Jual saja semuanya pada asing biar IKN Jadi sewa sama asing Ambisi sih boleh jika ngutang terus ya bangkrut wwwooooooY
137	Menembus Terowongan Multi Utilitas IKN Rapi Dan Dilapisi Epoksi di Bali Geby PO Bus SMK Lingga Kencana Depok OBnya Subang
138	Tanggal April ke IKN oleh pegawai proyek tdk dibolehkan masuk ke lokasi istana dan perumahan pejabat Krn kondisi dlm tdk memungkinkan utk dilalui mobil yg kami gunakan
139	Presiden Berkantor Di IKN Juli Smart Defense System Diterapkan di Bali Geby PO Bus SMK Lingga Kencana Dep OBnya
140	Pembangunan IKN Dan Visi Masa Depan Bangsa Dalam Model Dan Paradigma Komunikasi Pembangunan di Bali Geby PO Bus OBnya Tenggo Casemiro
141	Dari awal dah di kritik IKN ngak urgent Cm mau ambil alih Jakarta
142	ya bayar buat alasan bayar ganti tanah adatmu itu yg harganya setinggi langit bandingin sama IKN ya ga apel to apel lagian lucu juga sumbar ini disaat semua provinsi di sumatra sudah ada tol cuman dia yang menolak pal
143	Saya berharap semoga IKN mangkrak seperti Hambalang pepo
144	Uang pesongan jokowi dibelikan tanah di IKN sj biar jd pioneer warga disana
145	Iikut forum diskusi tentang ikn salpok sama bapak pemateri satu ini usia beliau tahun dan modis gaul bangett plisss mana mobil nya jazz warna kuning plus bumper belakang yg dimodif
146	Iyo Bapakmu ajak en ngarit nang IKN ngarit kuwi ngirit ra sah nganggo APBN
147	Sindiran Pedas Rocky Gerung Ke Jokowi Singgung Proyek IKN
148	Pejabat yg ikhlas dan kemungkinan betah tinggal di IKN ya Yang Mulya Jokodg keluarga Soloraya nya Kan mau Agustusan di IKN sekalian boyong semua keluarga ke sana sebagai contoh &
149	Suruh jadi gubernur IKN yang dibanggai sm org orgnga aja sih jadi wakilnya si ngesang
150	kalo mo ke ikn kudu lewat kota ini
151	Kalau presiden yg menggagas gemana dia balik solo apa ke IKN wakakakaka
152	Padahal IKN dibangun dari uang rakyat Sebenarnya IKN untuk rakyat atau pejabat dan elit ekonomi dari negara seberang ya
153	Makan tuh IKN
154	dah jelas kalo akan ada rencana pelantikan bakalan di lakuin di IKN Kok bisa nya bikin isu yang ga relevan

Dataset lengkap yang mencakup semua data yang digunakan dalam penelitian ini dapat diakses secara online melalui GitHub. Untuk mengakses dataset lengkap, silakan kunjungi repositori GitHub berikut:

Repository GitHub: <https://github.com/ayypcode/sentiment-analisist-ikn>

Dataset lengkap ini mencakup data-data yang diperlukan untuk mereplikasi eksperimen yang dilakukan dalam penelitian ini. Informasi lebih lanjut mengenai struktur dataset dan penjelasan mengenai atribut-atribut yang digunakan dapat ditemukan di dalam repositori tersebut.

CODE PROGRAM

```
import numpy as np
from sklearn.datasets import fetch_20newsgroups
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score
from pyswarm import pso
import pandas as pd
df_data = pd.read_csv('test.csv')
import re
def cleaningulasan(ulasan):
    ulasan = re.sub(r'@\w+', ' ', ulasan)
    ulasan = re.sub(r'#[A-Za-z0-9]+', ' ', ulasan)
    ulasan = re.sub(r"http\S+", ' ', ulasan)
    ulasan = re.sub(r'[0-9]+', ' ', ulasan)
    ulasan = re.sub(r"[-()#/@;:<>{}+=~|.!.?,_%]", " ", ulasan)
    ulasan = ulasan.strip(' ')
    return ulasan
df_data['Cleaning']= df_data['content'].apply(cleaningulasan)
def clearEmoji(ulasan):
    return ulasan.encode('ascii', 'ignore').decode('ascii')
df_data['HapusEmoji']= df_data['Cleaning'].apply(clearEmoji)
def replaceTOM(ulasan):
    pola = re.compile(r'(.)\1{2,}', re.DOTALL)
    return pola.sub(r'\1', ulasan)
df_data['3/Lebih']= df_data['HapusEmoji'].apply(replaceTOM)

df_data_cleaned = df_data.drop_duplicates(subset=['3/Lebih'])
```

```

df_data_cleaned.to_csv('cleaned_data.csv', index=False)

def casefoldingText(ulasan):
    ulasan = ulasan.lower()
    return ulasan

# Membaca data dari CSV yang sudah dibersihkan sebelumnya
df_data = pd.read_csv('cleaned_data.csv')

# Melakukan case folding
df_data['CaseFolding'] = df_data['3/Lebih'].apply(casefoldingText)
df_data.to_csv('casefolding.csv', index=False)
df_data[['Cleaning', 'CaseFolding']]

import nltk

from nltk.tokenize import word_tokenize
nltk.download('punkt')

# Membaca data dari CSV yang sudah melalui proses stopword removal
df_data = pd.read_csv('output_with_sentiment55.csv')

# Pastikan bahwa nilai dalam kolom Stemming adalah string
df_data['Normalization'] = df_data['Normalization'].astype(str)

# Tokenisasi
def tokenizingText(ulasan):
    ulasan = word_tokenize(ulasan)
    return ulasan

df_data['Tokenizing'] = df_data['Normalization'].apply(tokenizingText)
df_data.to_csv('token55.csv', index=False)

# Tampilkan hasil
df_data[['Normalization', 'Tokenizing']]

from nltk.corpus import stopwords

# Unduh stopwords NLTK jika belum diunduh
nltk.download('stopwords')

# Membaca data dari CSV yang sudah melalui proses stopword removal
df_data = pd.read_csv('token55.csv')

```

```

# Membuat daftar kata stop menggunakan NLTK untuk bahasa Indonesia
stop_words = set(stopwords.words('indonesian'))

# Fungsi menghapus stopwords

def removeStopWords(ulasan):
    words = ulasan.split()
    filtered_words = [word for word in words if word.lower() not in stop_words]
    return ''.join(filtered_words)

# Melakukan stopword removal
df_data['Stopword_Removal'] = df_data['Tokenizing'].apply(removeStopWords)

# Menyimpan hasil ke file CSV
df_data.to_csv('stopword_removal56.csv', index=False)

# Menampilkan hasil akhir
print(df_data[['Tokenizing', 'Stopword_Removal']])

import pandas as pd

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

# Membuat stemmer
factory = StemmerFactory()

stemmer = factory.create_stemmer()

# Membaca data dari CSV
df_data = pd.read_csv('stopword_removal55.csv')

# Fungsi untuk stemming

def stemmed_wrapper(term):
    return stemmer.stem(term)

# Membuat kamus untuk menyimpan hasil stemming
term_dict = {}

df_data['Stopword_Removal'] = df_data['Stopword_Removal'].fillna('')

# Mengisi kamus dengan kata-kata unik dari dokumen
for document in df_data['Stopword_Removal']:
    for term in document.split(): # Memastikan setiap term diolah

```

```

if term not in term_dict:
    term_dict[term] = ''

# Melakukan stemming pada setiap term
for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term, ":", term_dict[term])

# Fungsi untuk melakukan stemming pada teks
def stemmingText(document):
    return ''.join([term_dict[term] for term in document.split()])

# Menerapkan stemming pada kolom 'Stopword Removal'
df_data['Stemming'] = df_data['Stopword_Removal'].swifter.apply(stemmingText)

# Menyimpan hasil ke CSV
df_data.to_csv('stemming55.csv', index=False)

# Menampilkan dataframe
df_data

def convertToSlangword(ulasan):
    kamusSlang = eval(open("slangwords.txt").read())
    pattern = re.compile(r'\b( |'.join(kamusSlang.keys())+r')\b')
    content = []
    for kata in ulasan:
        filterSlang = pattern.sub(lambda x: kamusSlang[x.group()],kata)
        content.append(filterSlang.lower())
    ulasan = content
    return ulasan

df_data['Normalization'] = df_data['Stemming'].apply(convertToSlangword)

from googletrans import Translator
# Membuat instance dari Translator
translator = Translator()
# Membaca data dari file CSV yang berisi hasil normalisasi

```

```

df_data = pd.read_csv('normal55.csv')

# Fungsi untuk melakukan translasi teks

def translate_text(ulasan, src='id', dest='en'):

    try:
        # Melakukan translasi
        translation = translator.translate(ulasan, src=src, dest=dest)
        return translation.text
    except Exception as e:
        # Jika terjadi error, kembalikan teks asli
        print(f"Error during translation: {e}")
        return ulasan

# Terapkan fungsi translasi pada kolom 'Normalization_String'

df_data['Translated_Text'] = df_data['Normalization'].apply(lambda text:
translate_text(text))

# Menampilkan DataFrame untuk memverifikasi hasil translasi

df_data[['Normalization', 'Translated_Text']]

df_data.to_csv('Translated56_data.csv', index=False)

from textblob import TextBlob

# Definisikan fungsi analisis sentimen

def sentiment_analysis(text):

    blob = TextBlob(text)
    polarity = blob.sentiment.polarity
    if polarity > 0:
        return 'Positive'
    else:
        return 'Negative'

# Import data dari file CSV

file_path = 'Translated56_data.csv' # Ganti dengan path ke file CSV Anda

df_data = pd.read_csv(file_path)

# Terapkan analisis sentimen pada kolom 'Translated_Text'

```

```

df_data['Sentiment_Label'] = df_data['Translated_Text'].apply(sentiment_analysis)

# Menghitung jumlah sentimen positif dan negatif

positif_count = (df_data['Sentiment_Label'] == 'Positive').sum()
negatif_count = (df_data['Sentiment_Label'] == 'Negative').sum()

print("Jumlah Sentimen Positif:", positif_count)
print("Jumlah Sentimen Negatif:", negatif_count)

# Tampilkan DataFrame untuk memverifikasi hasil pelabelan
print(df_data)

# Simpan hasil ke file baru jika diperlukan
df_data.to_csv('output_with_sentiment14.csv', index=False)

import matplotlib.pyplot as plt

# Data dari df_data['sentimen'].value_counts()
sentimen_counts = {'Positif': 560, 'Negatif': 803}

# Membuat bar plot
plt.figure(figsize=(8, 6))

plt.bar(sentimen_counts.keys(), sentimen_counts.values(), color=['green', 'red'])

plt.title('Sentimen Komentar IKN')
plt.xlabel('Sentimen')
plt.ylabel('Jumlah')

plt.ylim(0, 1500) # Atur batas y agar lebih mudah melihat distribusi
plt.show()

# Vektorisasi Teks

vectorizer = TfidfVectorizer(max_features=5000)
X = vectorizer.fit_transform(df_data['Normalization'])

y = df_data['Sentiment_Label'].apply(lambda x: 1 if x == 'Positif' else 0)

from sklearn.feature_extraction.text import TfidfVectorizer

# Contoh data teks untuk vektorisasi
data = df_data['Stemming']

# Membuat DataFrame dari data
df_data = pd.DataFrame(data)

```

```

# Inisialisasi TfIdfVectorizer
vectorizer = TfIdfVectorizer(max_features=5000)

# Vektorisasi teks
X = vectorizer.fit_transform(df_data['Stemming']).toarray()

# Display terms with their overall rank based on the average TF-IDF score across all
documents

df_term_rank = pd.DataFrame(df_tfidf_example.mean()).reset_index()

df_term_rank.columns = ['Term', 'Average_TFIDF']

df_term_rank.sort_values(by='Average_TFIDF', ascending=False, inplace=True)

df_term_rank # Show the top terms with the highest average TF-IDF scores

from wordcloud import WordCloud

import matplotlib.pyplot as plt

plt.figure(figsize=(10, 5))

plt.imshow(wordcloud, interpolation='bilinear')

plt.axis('off') # Sembunyikan sumbu

plt.title('Word Cloud dari Skor TF-IDF')

plt.show()

# Memisahkan data menjadi training dan testing

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

from sklearn.naive_bayes import MultinomialNB

from sklearn.model_selection import cross_val_score, train_test_split

from sklearn.metrics import roc_auc_score, accuracy_score, roc_curve, auc,
confusion_matrix

from sklearn.datasets import load_digits

from sklearn.preprocessing import MinMaxScaler

from pyswarm import pso

import matplotlib.pyplot as plt

# Latih dan evaluasi model Naive Bayes tanpa optimasi

model_no_opt = MultinomialNB()

model_no_opt.fit(X_train, y_train)

y_pred_no_opt = model_no_opt.predict(X_test)

```

```

y_prob_no_opt = model_no_opt.predict_proba(X_test)

# Evaluasi kinerja model tanpa optimasi

accuracy_no_opt = accuracy_score(y_test, y_pred_no_opt)

fpr_no_opt, tpr_no_opt, _ = roc_curve(y_test, y_prob_no_opt[:, 1],
pos_label=model_no_opt.classes_[1])

roc_auc_no_opt = auc(fpr_no_opt, tpr_no_opt)

# Hitung confusion matrix untuk model tanpa optimasi

conf_matrix_no_opt = confusion_matrix(y_test, y_pred_no_opt)

print(f'Non-optimized Accuracy: {accuracy_no_opt}')

print(f'Non-optimized AUC: {roc_auc_no_opt}')

print("Confusion Matrix (No Optimization):")

print(conf_matrix_no_opt)

# Fungsi objektif yang mengembalikan negatif dari akurasi Naive Bayes

def nb_accuracy(alpha):

    model = MultinomialNB(alpha=alpha[0])

    scores = cross_val_score(model, X_train, y_train, cv=10, scoring='accuracy')

    return -np.mean(scores) # PSO meminimalkan fungsi, jadi kita balik tanda

# Batas atas dan bawah untuk alpha

lb = [0.01]

ub = [10]

# Optimasi menggunakan PSO

alpha_opt, fopt = pso(nb_accuracy, lb, ub, swarmsize=10, omega=1.5, phip=0.5, phig=0.5,
maxiter=50, minstep=1e-6)

print(f'Best alpha: {alpha_opt}')

# Latih model dengan alpha optimal dan mengevaluasi kinerja

model_opt = MultinomialNB(alpha=alpha_opt[0])

model_opt.fit(X_train, y_train)

y_pred_opt = model_opt.predict(X_test)

y_prob_opt = model_opt.predict_proba(X_test)

# Evaluasi kinerja model dengan alpha optimal

```

```

accuracy_opt = accuracy_score(y_test, y_pred_opt)

fpr_opt, tpr_opt, _ = roc_curve(y_test, y_prob_opt[:, 1], pos_label=model_opt.classes_[1])

roc_auc_opt = auc(fpr_opt, tpr_opt)

# Hitung confusion matrix untuk model dengan optimasi

conf_matrix_opt = confusion_matrix(y_test, y_pred_opt)

print(f'Optimal Accuracy: {accuracy_opt}')

print(f'Optimal AUC: {roc_auc_opt}')

print("Confusion Matrix (With PSO):")

print(conf_matrix_opt)

# Plot kurva ROC untuk kedua model

plt.figure(figsize=(10, 6))

plt.plot(fpr_no_opt, tpr_no_opt, color='blue', label=f'No PSO (AUC = {roc_auc_no_opt:.2f})')

plt.plot(fpr_opt, tpr_opt, color='green', label=f'With PSO (AUC = {roc_auc_opt:.2f})')

plt.plot([0, 1], [0, 1], color='red', linestyle='--')

plt.xlabel('False Positive Rate')

plt.ylabel('True Positive Rate')

plt.title('ROC Curve')

plt.legend(loc='lower right')

plt.show()

# Plot akurasi

plt.figure(figsize=(8, 5))

labels = ['No PSO', 'With PSO']

accuracies = [accuracy_no_opt, accuracy_opt]

plt.bar(labels, accuracies, color=['blue', 'green'])

plt.ylabel('Accuracy')

plt.title('Accuracy Comparison')

plt.ylim([0, 1])

plt.show()

from sklearn.metrics import confusion_matrix, ConfusionMatrixDisplay, accuracy_score

```

```

import matplotlib.pyplot as plt

# Confusion matrices

cm_nb = confusion_matrix(y_test, y_pred_no_opt)
cm_pso = confusion_matrix(y_test, y_pred_opt)

# Calculate accuracy values

accuracy_nb = accuracy_score(y_test, y_pred_no_opt)
accuracy_pso = accuracy_score(y_test, y_pred_opt)

# Plotting the confusion matrices

fig, axes = plt.subplots(1, 2, figsize=(12, 5))

# Naive Bayes without PSO

disp_nb = ConfusionMatrixDisplay(confusion_matrix=cm_nb, display_labels=['Negatif', 'Positif'])

disp_nb.plot(ax=axes[0], cmap=plt.cm.Blues, values_format='d')

axes[0].set_title(f'Confusion Matrix: Naive Bayes without PSO\nAccuracy: {accuracy_nb:.2f}')

# Naive Bayes with PSO

disp_pso = ConfusionMatrixDisplay(confusion_matrix=cm_pso, display_labels=['Negatif', 'Positif'])

disp_pso.plot(ax=axes[1], cmap=plt.cm.Blues, values_format='d')

axes[1].set_title(f'Confusion Matrix: Naive Bayes with PSO\nAccuracy: {accuracy_pso:.2f}')

plt.tight_layout()

plt.show()

# Accuracy values for plotting

methods = ['Naive Bayes\nwithout PSO', 'Naive Bayes\nwith PSO']

accuracies = [accuracy_nb, accuracy_pso]

# Define colors

colors = ['#1f77b4', '#ff7f0e']

# Create a bar chart

plt.figure(figsize=(10, 6))

bars = plt.bar(methods, accuracies, color=colors, edgecolor='black', linewidth=1.5)

# Adding accuracy values above bars

```

```

for bar in bars:
    yval = bar.get_height()

    plt.text(bar.get_x() + bar.get_width()/2.0, yval + 0.005, f'{yval:.2%}', ha='center',
             va='bottom', fontsize=12, fontweight='bold')

# Set the y-axis range from 60% to 80%
plt.ylim(0.6, 0.72)

# Adding labels and title
plt.ylabel('Accuracy', fontsize=14, fontweight='bold')

plt.title('Comparison of Accuracy\nNaive Bayes without PSO vs Naive Bayes with PSO',
          fontsize=16, fontweight='bold')

# Adding gridlines
plt.grid(axis='y', linestyle='--', alpha=0.7)

# Adding a background color to the plot
plt.gca().set_facecolor('#f5f5f5')

# Customizing ticks
plt.xticks(fontsize=12, fontweight='bold')
plt.yticks(np.arange(0.6, 0.7, 0.02), fontsize=12, fontweight='bold')

# Adding a legend
plt.legend(bars, ['Naive Bayes without PSO', 'Naive Bayes with PSO'], loc='upper left',
           fontsize=12)

# Display the plot
plt.tight_layout()
plt.show()

```



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
LEMBAGA PENELITIAN

Kampus Unisan Gorontalo Lt.3 - Jln. Achmad Nadjamuddin No. 17 Kota Gorontalo
Telp: (0435) 8724466, 829975 E-Mail: lembagapenelitian@unisan.ac.id

Nomor : 4762/PIP/LEMLIT-UNISAN/GTO/X/2023

Lampiran : -

Hal : Permohonan Izin Penelitian

Kepada Yth,

Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo
di,-

Tempat

Yang bertanda tangan di bawah ini :

Nama : Dr. Rahmisyari, ST.,SE.,MM
NIDN : 0929117202
Jabatan : Ketua Lembaga Penelitian

Meminta kesediannya untuk memberikan izin pengambilan data dalam rangka penyusunan **Proposal / Skripsi**, kepada :

Nama Mahasiswa : Arif Nur Alam
NIM : T3119112
Fakultas : Fakultas Ilmu Komputer
Program Studi : Teknik Informatika
Lokasi Penelitian : PLATFORM ONLINE TWITTER
Judul Penelitian : ANALISIS SENTIMEN TERHADAP CHATGPT
MENGGUNAKAN ALGORITMA SUPPORT VECTOR
MACHINE (SVM) PADA PLATFORM TWITTER

Atas kebijakan dan kerja samanya diucapkan banyak terima kasih.



+



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS IHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001
Jl. Achmad Nadjamuddin No. 17 Telp (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT KETERANGAN PENELITIAN

Nomor : 102 /FIKOM-UIG/SKP/V/2024

Yang bertanda tangan dibawah ini :

Nama : Irvan Abraham Salihi, M. Kom
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini Menerangkan bahwa :

Nama Mahasiswa : Arif Nur Alam
NIM : T3119112
Program Studi : Teknik Informatika

Bahwa yang bersangkutan benar-benar telah melakukan penelitian tentang "**Optimasi Naïve Bayes Berbasis PSO Untuk Analisis Sentimen Terhadap Pembangunan Ibu Kota Nusantara Pada Twitter**" Guna untuk meyelesaikan Studi pada Program Studi Teknik Informatika Fakultas Ilmu Komputer, dan bersangkutan telah menyelesaikan penelitian tersebut pada **TGL 20 Mei** sesuai dengan waktu yang telah ditentukan.

Demikian Surat Keterangan ini dibuat dan digunakan untuk seperlunya.

Gorontalo, 20 Mei 2024
Dekan,

Irvan A. Salihi, S.Kom, M.Kom
NIDN : 0928028101



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS IHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
UPT. PERPUSTAKAAN FAKULTAS
SK. MENDIKNAS RI NO. 84/D/0/2001
Jl. Achmad Nadjamuddin No.17 Telp(0435) 829975 Fax. (0435) 829976 Gorontalo

SURAT KETERANGAN BEBAS PUSTAKA

No : 013/Perpustakaan-Fikom/VI/2024

Perpustakaan Fakultas Ilmu Komputer (FIKOM) Universitas Ihsan Gorontalo dengan ini menerangkan bahwa :

Nama Anggota : Arif Nur Alam

No. Induk : T3119112

No. Anggota : M202433

Terhitung mulai hari, tanggal : Jumat, 07 Juni 2024, dinyatakan telah bebas pinjam buku dan koleksi perpustakaan lainnya.

Demikian keterangan ini di buat untuk di pergunakan sebagaimana mestinya.



Gorontalo, 07 Juni 2024

**Mengetahui,
Kepala Perpustakaan**



Apriyanto Alhamad, M.Kom

NIDN : 0924048601



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS ICHSAN GORONTALO
FAKULTAS ILMU KOMPUTER
SURAT KEPUTUSAN MENDIKNAS RI NOMOR 84/D/O/2001
Jl. Achmad Najamuddin No. 17 Telp. (0435) 829975 Fax (0435) 829976 Gorontalo

SURAT REKOMENDASI BEBAS PLAGIASI
No. 124/FIKOM-UIG/R/VI/2024

Yang bertanda tangan di bawah ini :

Nama : Irvan Abraham Salihi, M.Kom
NIDN : 0928028101
Jabatan : Dekan Fakultas Ilmu Komputer

Dengan ini menerangkan bahwa :

Nama Mahasiswa : Arif Nur Alam
NIM : T3119112
Program Studi : Teknik Informatika (S1)
Fakultas : Fakultas Ilmu Komputer
Judul Skripsi : Optimasi Naive Bayes Berbasis PSO Untuk Analisis Sentimen Terhadap Pembangunan Ibu Kota Nusantara Pada Twitter

Sesuai hasil pengecekan tingkat kemiripan skripsi melalui aplikasi Turnitin untuk judul skripsi di atas diperoleh hasil Similarity sebesar 16%, berdasarkan Peraturan Rektor №. 32 Tahun 2019 tentang Pendekripsi Plagiat pada Setiap Karya Ilmiah di Lingkungan Universitas Ichsan Gorontalo dan persyaratan pemberian surat rekomendasi verifikasi calon wisudawan dari LLDIKTI Wil. XVI, bahwa batas kemiripan skripsi maksimal 30%, untuk itu skripsi tersebut di atas dinyatakan BEBAS PLAGIASI dan layak untuk diujangkan.

Demikian surat rekomendasi ini dibuat untuk digunakan sebagaimana mestinya.

Mengetahui
Dekan,

Irvan Abraham Salihi, M.Kom
NIDN. 0928028101

Gorontalo, 08 Juni 2024
Tim Verifikasi,

Zulfrianto Lamasiqi, M.Kom
NIDN. 0942089101

Terlampir :
Hasil Pengecekan Turnitin

PAPER NAME	AUTHOR
SKRIPSI_T3119112_ARIF_NUR_ALAM.pdf	ARIF NUR ALAM arifnuralam101@gmail.com
WORD COUNT	CHARACTER COUNT
9802 Words	61665 Characters
PAGE COUNT	FILE SIZE
59 Pages	2.6MB
SUBMISSION DATE	REPORT DATE
Jun 8, 2024 8:06 AM GMT+8	Jun 8, 2024 8:07 AM GMT+8

● 16% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 15% Internet database
- Crossref database
- 0% Submitted Works database
- 5% Publications database
- Crossref Posted Content database

● Excluded from Similarity Report

- Bibliographic material
- Cited material
- Quoted material
- Small Matches (Less than 10 words)

● 16% Overall Similarity

Top sources found in the following databases:

- 15% Internet database
- Crossref database
- 0% Submitted Works database
- 5% Publications database
- Crossref Posted Content database

TOP SOURCES

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	senafti.budiluhur.ac.id	2%
	Internet	
2	e-journal.stmiklombok.ac.id	2%
	Internet	
3	ejurnal.unisan.ac.id	1%
	Internet	
4	repository.uir.ac.id	1%
	Internet	
5	eprints.walisongo.ac.id	<1%
	Internet	
6	etheses.uin-malang.ac.id	<1%
	Internet	
7	ojs.trigunadharma.ac.id	<1%
	Internet	
8	jacis.pubmedia.id	<1%
	Internet	

9	inews.id Internet	<1%
10	lkn.kompas.com Internet	<1%
11	repository.ittelkom-pwt.ac.id Internet	<1%
12	researchgate.net Internet	<1%
13	eprints.utdi.ac.id Internet	<1%
14	es.scribd.com Internet	<1%
15	repository.nusaputra.ac.id Internet	<1%
16	repository.uin-suska.ac.id Internet	<1%
17	repository.ub.ac.id Internet	<1%
18	ejournal-binainsani.ac.id Internet	<1%
19	Moh. Aulia Miftakhurahmat, Nur Safitri, Putri Aulia Kusnadi, Chaerur Ro... Crossref	<1%
20	repository.nobel.ac.id Internet	<1%

[Sources overview](#)

21	dspace.uii.ac.id Internet	<1%
22	Taghfirul Azhima Yoga Siswa, Renaldi Panji Wibowo. "Komparasi Meto... Crossref	<1%
23	docplayer.info Internet	<1%
24	ejurnal.umri.ac.id Internet	<1%
25	neliti.com Internet	<1%
26	Budi Kurniawan, Achmad Suwarisman, Iis Afriyanti, Aditya Wahyudi, De... Crossref	<1%
27	Dedy Atmajaya, Annisa Febrianti, Herdianti Darwis. "Metode SVM dan ... Crossref	<1%
28	ejournal.nusamandiri.ac.id Internet	<1%
29	123dok.com Internet	<1%
30	merahputih.com Internet	<1%
31	ejournal.itn.ac.id Internet	<1%
32	repository.its.ac.id Internet	<1%

[Sources overview](#)

33	msn.com Internet	<1%
34	Irma Surya Kumala Idris, Yasin Aril Mustofa, Irwan Abraham Salihi. "An... Crossref	<1%
35	ejournal.unsri.ac.id Internet	<1%
36	ejurnal.ars.ac.id Internet	<1%
37	stat.ipb.ac.id Internet	<1%
38	Ahmad Santoso, Agung Nugroho, Aswan S Sungi. "Analisis Sentimen ... Crossref	<1%
39	Wiyanto Wiyanto, Zulita Setyaningsih. "Sentiment Analysis Pemutusan ... Crossref	<1%
40	repo.palcomtech.ac.id Internet	<1%
41	repository.stei.ac.id Internet	<1%

Sources overview

Daftar Riwayat Hidup



Nama	:	Arif Nur Alam
Nim	:	T3119112
Tempat, Tanggal Lahir	:	Kupang, 13 Januari 1999
Alamat	:	Desa Talango
Agama	:	Islam
Kewarganegaraan	:	WNI
Email	:	arifnuralam101@gmail.com

Riwayat pendidikan dan pekerjaan

1. Tahun 2011, Menyelesaikan Pendidikan di Sekolah Dasar Negeri 11 Kabilia
2. Tahun 2014, Menyelesaikan Pendidikan di MTs Al-Khairaat Kota Gorontalo
3. Tahun 2017, Menyelesaikan Pendidikan di MAN 1 Kota Gorontalo
4. Tahun 2019, Diterima menjadi mahasiswa di Perguruan Tinggi Swasta Universitas Ichsan Gorontalo